DOCUMENT RESUME

ED 245 293 CS 504 630

AUTHOR	Rubin, Donald L.; Mead, Nancy A.
TITLE	Large Scale Assessment of Oral Communication Skills: Kindergarten through Grade 12.
INSTITUTION	ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill.; Speech Communication Association, Annandale, Va.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	84
CONTRACT	400-82-0025
NOTE	130p.
AVAILABLE FROM	Speech Communication Association, 5105 Backlick Rd., Suite E, Annandale, VA 22003 ($11.95; 10% discount for members).
PUB TYPE	Guides - Non-Classroom Use (055) -- (13Information Analyses - ERIC Information Analysis Produc
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Academic Standards; Educational Assessment; Elementary Secondary Education; Evaluation Criteria; *Language Tests; *Listening Skills; *Oral Language; *Speech Communication; *Speech Skills; *Speech Tests; Student Evaluation; Testing; Test Reviews

ABSTRACT

Intended for teachers and administrators who wish to assess student oral communication needs before designing an appropriate program, this guide provides a review of tests for measuring acts of speaking and listening. The guide surveys and discusses procedures for assessing speaking and listening skills among school children and focuses on technical issues of measurement and pragmatic questions of administrative feasibility. The first section provides a review and critique of procedures for assessing oral communication skills. The second section reviews 45 oral communication assessment instruments, including the California Achievement Test for Listening, Comprehensive Test of Basic Skills, Metropolitan Achievement Tests, and the National Assessment of Educational Progress Pilot Test of Speaking and Listening. Appendixes contain standards for effective oral communication programs, and criteria for evaluating instruments and procedures for assessing speaking and listening. (HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED245293

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

LARGE SCALE ASSESSMENT OF ORAL COMMUNICATION SKILLS:

Kindergarten through Grade 12

Don L. Rubin
Nancy A. Mead

CS 504 630

Large Scale Assessment of Oral Communication Skills: Kindergarten through Grade 12

Donald L. Rubin
The University of Georgia

Nancy A. Mead
Education Commission of the States

in collaboration with
John Daly, University of Texas, Austin
Patrick Dickson, University of Wisconsin, Madison
James C. McCroskey, West Virginia University, Morgantown
Janis Patterson, University of Wisconsin, Madison

ERIC Clearinghouse on Reading and Communication Skills
National Institute of Education

Speech Communication Association
5105 Backlick Road, Annandale, Virginia 22003

Speech Communication Association Publication Board: Gustav W. Friedrich, University of Oklahoma, *Chairperson*; Vicki S. Freimuth, University of Maryland; Robert L. Scott, University of Minnesota; William Work, Speech Communication Association, *ex officio*.

Consultant Readers: Janis F. Andersen, San Diego State University; Philip M. Backlund, Central Washington University; Kenneth L. Brown, University of Massachusetts.

Staff Editor: Barbara Davis

Cover Design: Joyce Conley

Library of Congress Catalog Card Number: 84-50269

Published 1984 by the ERIC Clearinghouse on Reading and Communication Skills, 1111 Kenyon Road, Urbana, Illinois 61801, and the Speech Communication Association, 5105 Backlick Road, Annandale, Virginia 22003. Printed in the United States of America.

This publication was prepared with funding from the National Institute of Education, U.S. Department of Education, under contract no. 400-82-0025. Contractors undertaking such projects under government sponsorship are encouarged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to the Speech Communication Association for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either the Speech Communication Association or the National Institute of Education.

Preface

Because the evaluation of oral communication is an integral part of the education process, the Speech Communication Association is publishing this review of tests for measuring acts of speaking and listening. Teachers and administrators, desiring to undertake that all-important first step of assessing student needs before a program is designed, ask, "Which test should I use?" *This guide provides direction as to what choices are available*. Since states and school districts vary in their program goals for speaking and listening, the choice of tests will also vary.

Realizing educators might want a professional evaluation of currently available tests, a special Committee on Assessment and Instrument Development was appointed by the Speech Communication Association's Committee on Assessment and Testing. Don Rubin and Nancy Mead were major participants on that committee; their thorough work led to this publication.

Whether a state evaluation is taking place or a teacher wants a test for the classroom, this review should be helpful. Test validity and reliability are common concerns for the educator constructing tests; equally important are questions of cost, time, and scoring procedures. The forty-five tests reviewed in Chapter II include each of these concerns, as well as a full description of the assessment instrument. To help in the search for comparable tests, the age range and skills tested are listed first. Also included is a section on evaluative comments. These professional judgments should provide the reader with information necessary to make good choices.

Testing is a value laden subject. For some, all testing is loathsome and antithetical to the spirit of learning. Others think testing is burdensome but necessary for providing feedback about educational outcomes. Yet another point of view regards evaluation opportunities—as a powerful tool for shaping curriculum or manipulating educational policy.

This review surveys and discusses procedures for assessing speaking and listening skills among public school aged children. In doing so, it focuses on technical issues of measurement and pragmatic questions of administrative feasibility. But the materials also emphasize the responsibility in testing, and users of this document must not lose sight of the purposes of oral communication assessment and weigh the variety of costs against the potential value of test results. For communication does not thrive in an exclusively evaluative climate. Assessment is beneficial only to the degree that it yields interpretable results—data that are directed toward solving educational problems, data that reflect communication skills and behaviors that are central to effective functioning; not merely data that is readily measurable. In this light, the task of evaluating, selecting, or developing appropriate measurement instruments demands that educators render decisions on as informed a basis as possible. It is our hope that this information will facilitate such decisions, and thus contribute to responsible evaluation.

In any testing situation the user must know what is and what is not being tested. In any testing situation the user must know how to communicate the results of the test *with the qualifications* inherent in the testing situation. The purpose of the information provided in this guide is to aid in that communication.

Foreword

The Educational Resources Information Center (ERIC) is a national information system developed by the U.S. Office of Education and now sponsored by the National Institute of Education (NIE). It provides ready access to descriptions of exemplary programs, research and development efforts, and related information useful in developing more effective educational programs.

Through its network of specialized centers or clearinghouses, each of which is responsible for a particular educational area, ERIC acquires, evaluates, abstracts, and indexes current significant information and lists this information in its reference publications.

ERIC/RCS, the ERIC Clearinghouse on Reading and Communication Skills, disseminates educational information related to research, instruction, and personnel preparation at all levels and in all institutions. The scope of interest of the Clearinghouse includes relevant research reports, literature reviews, curriculum guides and descriptions, conference papers, project or program reviews, and other print materials related to all aspects of reading, English, educational journalism, and speech communication.

The ERIC system has already made available—through the ERIC Document Reproduction System—much informative data. However, if the findings of specific educational research are to be intelligible to teachers and applicable to teaching, considerable bodies of data must be reevaluated, focused, translated, and molded into an essentially different context. Rather than resting at the point of making research reports readily accessible, NIE has directed the separate clearinghouses to work with professional organizations in developing information analysis papers in specific areas within the scope of the clearinghouses.

ERIC is pleased to cooperate with the Speech Communication Association in making *Large Scale Assessment of Oral Communication Skills: Kindergarten through Grade 12* available.

Charles Suhor
Director, ERIC/RCS

Contents

I. A Review and Critique of Procedures for Assessing Oral Communication Skills 1

Why Teach Oral Communication Skills? 1
Background and Objectives 3
Procedures for Reviewing Assessment Instruments 4
Types of Response and Scoring Procedures 14
Content of Assessment Instruments 15
Administrative Feasibility 17
Target Populations and Potential Sources of Test Bias 18
Locally Developed Instruments 19
Selected Research and Development Priorities 21

II. Reviews of Oral Communication Assessment Instruments 27

Appendix A Standards for Effective Oral Communication Programs 119
Appendix B Critera for Evaluating Instruments and Procedures for Assessing Speaking and Listening 120

Bibliography 122

I. A Review and Critique of Procedures for Assessing Oral Communication Skills

WHY TEACH ORAL COMMUNICATION SKILLS?[1]

A well-known adage has it that of all the creatures inhabiting the earth, fish are the least likely to ever discover water. So it is with communication. Speech comes to us as part of our innate endowment as human beings. We are enguifed by communication in all our daily affairs. Usually we are not directly aware of our oral communication environment. But it is nonetheless vital to our well-being and survival.

Speaking and listening are prerequisites to success in school. Most instructions for classroom procedures are delivered orally by teachers. Consequently, students with deficient listening skills fail to absorb much of the material to which they are exposed. Their problems are intensified when they respond incorrectly because they do not listen to questions carefully. Students who listen poorly are often isolated and left out of classroom activities. Speech performance also affects academic achievement. Students who cannot adequately express their knowledge are judged ignorant. Some speech styles trigger stereotyped expectations of poor ability—expectations that are likely to be self-fulfilling (Williams, Whitehead, and Miller, 1972). Quiet children may be appreciated for their "good behavior," but they are subject to similarly negative school experiences (McCroskey and Daly, 1976). Students who do not ask for assistance will not receive adequate assistance. One research study, for example, found that reticent students progressed slowly through a self-paced reading program, despite normal levels of reading aptitude. The reason for their poor performance was that these students rarely approached teachers for help (Scott, Yates, and Wheeless, 1975).

Beyond the confines of school, oral communication proficiency contributes to social adjustment and satisfying interpersonal relationships. Youngsters with poor communication skills are sometimes viewed as unattractive by their peers and enjoy few friendship bonds (Hurt and Preiss, 1978). Antisocial and violent behavior is frequently attributable to underdeveloped social sensitivity and lack of conflict resolution techniques. Remediation programs have reduced the incidence of antisocial acts by means of communication training (Chandler, 1973). Counselors acknowledge that many family problems are caused by poor communication, and may be ameliorated by improving interaction between family members (Shure and Spivack, 1978).

Speaking and listening are no less crucial in the marketplace. Communication skills rank high among lists of managerial competencies. An officer of one computer firm, for example, states that the company prefers to conduct its own training in computer programming, but seeks employees with strong communication abilities (Gruner, Logue, Freshley, and Huseman, 1977). Professionals—doctors, lawyers, engineers, teachers—require more than just subject matter expertise. These professionals must listen effectively to their patients, clients, or students in order to identify and analyze problems. They must speak effectively in order to implement their solutions. Individuals who speak in a nonstandard fashion (DeLaZerda and Hopper, 1978)

---

[1]An earlier version of this section appeared in D. Rubin and R. Bazzle, *Development of an Oral Communication Assessment Program. The Glynn County Speech Proficiency Examination for High School Students* (Brunswick, Georgia: Glynn County School System, 1981). The authors express their appreciation to the Glynn County School System for use of this material.

or who withdraw from speaking (Richmond, 1977) tend to be regarded by personnel officers as prospects for only low status, low paying jobs. Nonetheless, even unskilled workers have occasion to engage in job related speech, including a surprising amount of public speaking (Kendall, 1974).

Speech curricula have traditionally stressed the importance of communication for the preservation of a democratic society. Throughout its history, America has fought vigorously to safeguard freedom of expression under the assumption that full citizen participation is the surest guarantee against tyranny. Surely not every citizen deliberates as a member of a legislative body, but numerous opportunities for citizen input are available. These include participation in civic associations, public hearings, and citizen lobbying. At the very least, citizens are responsible for staying informed, and much of the pertinent information must be culled by listening.

Finally, oral communication is essential to full psychological development. Self-concept is acquired through interaction with others (Mead, 1934). Self-actualization, a sense of fulfillment (Maslow, 1954), usually involves interpersonal activities—making contributions, exerting influence, or being recognized in a social manner. In addition, speech is a means for artistic expression and self-discovery.

The fact that all students come to school with basic speaking and listening skills and also seem to develop more mature behaviors on their own as they grow older does not imply that all students are *effective* communicators. Educators occasionally comment, "My students don't need to learn how to talk. That's one thing they do too much of." But effective communication must be cultivated. Students may lack clarity in their speech. Their listening comprehension may not attain its fullest potential. Students who communicate well in familiar settings may lack the confidence and flexibility needed to express themselves effectively in a wider range of situations. Educators cannot rely on haphazard, unguided learning outside of the classroom to impart communication effectiveness. Systematic instruction is imperative.

Still, of all the basic skills, speaking and listening are most often neglected in schools. This neglect transpires despite numerous curriculum documents that urge attention to oral abilities. Undoubtedly, a host of factors discourage teachers from implementing oral communication instruction. Teachers are held accountable for students' reading achievement, for performance on mandated grammar tasks, for monitoring attendance, for giving enough homework, for not giving too much homework. But teachers are generally not held accountable for teaching students to speak and listen effectively. Furthermore, few teachers have received training in communication education or have materials available to aid instruction. Consequently, little concerted instruction in speech communication takes place.

If students' speaking and listening proficiency were systematically evaluated, it is likely that schools would systematically implement oral communication instruction. One substantial benefit of large scale assessment of oral communication skills is that such testing can guide innovation in this curriculum domain. Indeed, experience in Great Britain and elsewhere demonstrates that speech assessment has a "washback" effect on the amount and kinds of speech teaching undertaken in classrooms (Barnes, 1980).

Another benefit of oral communication assessment is that test results can be used to make decisions about the best manner in which to place individual students in instructional sequences. Assessment procedures that yield fine-grained analyses, rather than global judgments, can be used for diagnostic purposes (Rubin, 1981). Thus, for example, students who have difficulty in vocal production factors might concentrate on oral reading, while those whose difficulties lie in the area of organization might cycle through a set of story-telling exercises before progressing to explanatory discourse. Students who demonstrate strengths in, say, literal comprehension of spoken materials might advance to instructional units emphasizing critical listening skills.

Speaking and listening tests can also provide valuable information for program evaluation. Since large scale programs of oral communication improvement are in their infancy, it is especially important to evaluate their effectiveness and to secure data that will enable these programs to be "fine tuned." Program (and teacher) effectiveness is best judged with reference to student achievement on program objectives. If students are not achieving criterion performance levels in language use, for example, teachers and administrators will recognize that additional instructional effort needs to be directed to this area. It is worth noting, however, that student achievement can be interpreted as an indicator of program success only when student aptitude and institutional resources—the raw materials with which the program has to work—are also taken into account. Also, student achievement is not the only data that might contribute to program evaluation. Attitudinal outcomes, self-, and peer-evaluations are also useful information for this purpose. After emphasis on listening, the listening scores might not improve as much as teacher reports on better responses to directions on assignments.

A final use for speaking and listening assessments is to certify students as having attained (or not attained) mastery in oral communication. Certification in basic skills is increasingly demanded by competency-based education movements. Promotion or graduation decisions may be based upon such certification. Backlund and his associates (1981) surveyed many of the state and local jurisdictions that have already adopted large scale tests of speaking and listening skills as a means of certifying students' competencies in the basic skills.

BACKGROUND AND OBJECTIVES

Several developments in the past few years motivated the test review effort represented in this publication. Set in the background of a nationwide movement toward competency-based education, speech communication educators recognized that their field was not exempt from the challenge of accounting for educational outcomes (Ritter, 1978). A major step in facilitating that accountability was the publication of the results of the National Speech Communication Competencies Project (Allen and Brown, 1976). This document examined the development of communication skills and promulgated a description of functional communication competence that has helped guide many subsequent efforts in curriculum, instruction, and evaluation. A list of competencies in speaking and listening for high school graduates (Bassett, Whittington, and Staton-Spicer, 1978) has likewise influenced teaching and assessment. In their "Standards for Effective Oral Communication Programs" (see Appendix A), the American Speech-Language-Hearing Association and the Speech Communication Association asserted that effective instructional efforts must include provisions for appropriate and constructive methods of assessment and evaluation. Such methods were further clarified by "Criteria for Evaluating Instruments and Procedures for Assessing Speaking and Listening" (see Appendix B), a document endorsed by the Speech Communication Association.

Despite this initial impetus to evaluate communication competencies, despite the view that developing assessment procedures presents no insurmountable technical obstacles (Larson, 1978; McGlone, 1973), and despite some concrete suggestions of pertinent measurement instruments (Larson, Backlund, Redmond, and Barbour, 1978; McCaleb, 1979; McCaleb and Korman, 1978), attempts to implement large scale assessments of speaking and listening skills have not been forthcoming. In general, evaluation programs have been stymied by a scarcity of suitable instruments (Brown, Backlund, Gurney, and Jandt, 1979; Plattor, Unruh, Muir, and Loose, 1978). Consequently, the Steering Committee of the Task Force on Assessment and Testing of the Speech Communication Association acted to establish committees for the purpose of identifying existing instruments and furthering the development of additional instruments

for the measurement of communication skills. This monograph is an outgrowth of one such committee.[2]

Four primary objectives guided this effort:

- To monitor existing assessment instruments in oral communication.
- To abstract and describe assessment instruments and systematically report their availability to the Task Force and the Association.
- To study and recommend research that is needed to develop listening and speaking assessment instruments for elementary and secondary schools.
- To encourage development of new instruments by commercial and noncommercial sources.

In order to delimit the scope of the task and to set priorities, the committee defined its focus further. Particular attention was paid to assessment instruments that evaluate communication behaviors, as opposed to instruments that describe behaviors but assign no judgments of quality. In addition, emphasis was placed on measures of communication per se (verbal and nonverbal encoding and decoding in situations ranging from high interaction to extended and uninterrupted discourse), rather than on measures that focused exclusively on component subskills like language, role-taking, articulation, or perceptual acuity. The effort was directed toward instruments that had the measurement of communication as their main purpose, not those that used communication incidentally as a means to measure other skills. This latter emphasis did not necessarily exclude indirect measures of communication competency, but it severely constrained the types of indirect measures that might be found suitable. The search was narrowed to assessment procedures that seemed amenable to large scale testing in institutional/school settings. Finally, attempts were made to include instruments appropriate for a variety of individuals including non-native speakers, minority culture children, and students with special needs.

PROCEDURES FOR REVIEWING ASSESSMENT INSTRUMENTS

Given the objectives and emphases described in the preceding section, a number of potential sources of existing instruments were searched. Letters were sent to major commercial test publishers. Previously published compendiums of communication measures (Larson, et al., 1978; Brown, et al., 1979; Plattor, et al., 1978) were consulted, as were more general lists of tests and evaluation instruments (Fagan, Cooper, and Jensen, 1975; Buros, 1978; Johnson, 1976; Grommon, 1976). The literature on second language testing was also a valuable source of information (e.g., Lange and Clifford, 1980; Richards, 1981). The assessment committee collected a number of evaluation procedures produced by state and local education agencies. In addition, an ERIC search was conducted. Calls for assessment instruments were published in *SPECTRA* and in the *Newsletter of the National Conference for Research in English*. Finally, individual committee members contributed to the data base by examining literature in their areas of expertise.

A catalogue of instruments that met the criteria of the assessment committee is presented in Table 1. Each instrument was assigned nonsystematically to a single committee member for review. These reviews appear in Chapter 2 of this publication. The contents of the reviews reflect the views of the individual reviewers as influenced by their expert judgment.

The form used for the instrument reviews presented in this book is primarily descriptive.

---

[2]Other committee members who provided input to this report are J. Daly (University of Texas), W. P. Dickson (University of Wisconsin), and J. McCroskey (West Virginia University). Their contributions to this effort are gratefully acknowledged.

TABLE 1

CATALOGUE INSTRUMENTS REVIEWED

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
1	Brown-Carlson Listening Test	Harcourt, Brace & World, New York 10017	listening	high school, adult	administered orally and completed on standardized forms
2	California Achievement Test; Listening for Information; Level 10	CTB/McGraw-Hill, Del Monte Research Park, Monterey, CA 93940	listening	primary	group administered; multiple choice, paper and pencil format
3	CIRCUS Listening Test	Addison-Wesley, Reading, MA 01867	listening	grades k–3	group administered multiple-choice format
4	CIRCUS Say and Tell	Addison-Wesley, Reading, MA 01867	oral language	pre-k to 3	individual, oral; child responds to a variety of stimuli
5	Cloze Listening Test	John S. Bowdidge, 2017 S. Oak Grove Avenue, Springfield, MO 65804	listening	secondary	group administered; fill in the blank, paper and pencil format; tape recorded stimulus
6	Comprehensive Tests of Basic Skills, Tests 2, 3, and 4	CTB/McGraw Hill, Del Monte Research Park Monterey, CA 93940	listening; visual decoding; auditory discrimination	early elementary	group admistered; paper and pencil, multiple-choice format
7	Communicative Evaluation Chart from Infancy to Five Years	Educators Publishing Service, Cambridge, MA 02138	oral language; listening; social development; auditory perception	infancy–5 years	observer records presence or absence of skills on basis of extended observation
8	Durrell Listening-Reading Series: Primary, Intermediate, Advanced Levels	Harcourt Brace Jovanovich, Inc., New York, NY 10017	listening	grades 1–9	group administered; multiple-choice, paper and pencil format
9	Dyadic Task-Oriented Communication	C. A. Findley, ERIC Document Reproduction Service No. 145 629	speaking	elementary	administered to pairs of students, one presents task, other responds; responses tape recorded

TABLE 1 (cont.)

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
10	DYCOMM: Dyadic Communication	B.H. Byers, *DYCOMM: Dyadic Communication*, Honolulu: University of Hawaii, 1973	speaking; listening; interaction	adaptable k–12	groups of 10 or more students work in pairs rotating among partners and tasks
11	Fullerton Language Test for Adolescents	Consulting Psychologists Press, Palo Alto, CA 94306	listening; auditory processing	12–18 years; learning disabled and non-disabled	individual administration
12	Fundamental Achievement Series, Verbal	The Psychological Corporation, 757 Third Avenue, New York, NY 10017	listening; receptive language	grades 6–12	group administered; multiple-choice, paper and pencil format; taped instruction
13	Gary, Indiana Oral Proficiency Examination	Gary Community School Corporation; Gary, IN 46401	speaking	grade 10	individual speech performance addressed to examiner
14	Glynn County Speech Proficiency Examination	CBE Demonstration Project, Glynn County, Board of Education, Brunswick, GA 31521	speaking	secondary	simulated public hearing, students presenting arguments one at a time; responses videotaped
15	Language Assessment Scales	Linguametrics Group, P.O. Box 454, Corte Madera, CA 94925	speaking; listening	grades 1–5; Spanish or English	multiple-choice responses to oral presentations; oral imitation of sounds and words
16	Language Dominance Survey	Multilingual Center, Berkeley, California	speaking; listening	grades k–12; Spanish, English	individual administration
17	Language Facility Test	The Allington Corporation, 801 N. Pitt St., Alexandria, VA 22314	speaking	ages 3–15 for normal populations	individually administered; free responses to picture stimuli

TABLE 1 (cont.)

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
18	Language Skills Communication Task	M. C. Wang, S. Rose, & J. Maxwell, *The Development of the Language Communication Skills Test*. Pittsburgh: University of Pittsburgh Learning Research and Development Center, 1973	speaking; listening; interaction	k–?	students work in pairs; responses are recorded for subsequent scoring
19	Listening Comprehension Tests	A. Wilkinson, L. Stratta and P. Dudley, *Listening Comprehension Tests*. Macmillan Education Ltd., Houndsmills, Bassingstoke Hampshire, England RG21 2X5	listening	ages 10–11, 13–14, and 17–18	group administered; paper and pencil, multiple-choice format
20	MACOSA Listening and Speaking Tests	E. Plattor, W.R. Unruh, L. Muir & K.D. Loose *Test Development for Assessing Achievement in Listening and Speaking*. The Minister's Advisory Committee on Student Achievement Planning and Research, Alberta Education, 10105 109 Street, Edmonton, Alberta, Canada T5J 2V2	speaking; listening	grades 3, 6, 9, and 12	oral speaking test administered to small groups, each student responding in turn; responses tape recorded; written speaking test and listening test group administered; paper and pencil, multiple-choice format

TABLE 1 (cont.)

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
21	Massachusetts Assessment of Basic Skills Listening Test	Massachusetts Department of Education, Bureau of Research and Assessment, Boston, MA 02116	listening	grades 7–12	group administered with tape recorded instructions, listening passages, and multiple-choice response
22	Massachusetts Assessment of Basic Skills Speaking Test	Massachusetts Department of Education, Bureau of Research and Assessment, Boston, MA 02116	speaking	grades 7–12	two-tiered system with classroom teachers rating typical speaking abilities, and individual interviews for students who fail to pass the initial screening
23	Measure of Communication Competence	S. C. Riccillo, *Children's Speech and Communicative Competence*, Unpublished doctoral dissertation, University of Denver, 1974 University Microfilms No. 75-2210	speaking	ages 2½ to 4 years	individually administered, responses tape recorded
24	Metropolitan Achievement Tests: Listening Comprehension	The Psychological Corporation, 757 Third Avenue, New York, NY 10017	listening	grades k–4	group administered; multiple-choice, paper and pencil format
25	Michigan Educational Assessment Program: Listening Test	Michigan Educational Assessment Program, Michigan Department of Education, P.O. Box 30008, Lansing, MI 48909	listening	grades 4, 7, and 10	group administered; paper and pencil, multiple-choice format

TABLE 1 (cont.)

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
26	National Assessment of Educational Progress Pilot Test of Speaking and Listening	See, N.A. Mead, *The Development of an Instrument for Assessing Functional Communication Competence in Seventeen-Year-Olds*. Unpublished dissertation, University of Denver, 1977	speaking; listening; attitudes	age 17	group administered; multiple-choice, paper and pencil format; tape recorded instructions
27	New York State Regents Comprehensive Examination in English: Listening Section	Division of Educational Testing, New York State Education Department Albany, NY 12234	listening	grade 12	group administered; examiner reads passages aloud; multiple-choice format
28	New York Statewide Achievement Examination in English	Division of Educational Testing, New York State Education Department, Albany, NY 12234	speaking; listening	grade 12	for speaking section, students present brief monologues on supplied topics in class; listening section is group administered; passages are read aloud; multiple-choice format
29	Oliphant Tests: Auditory Synthesizing Test and Auditory Discrimination Memory Test	Educators Publishing Service; Cambridge, MA 02138	auditory memory	age 7–14	sounds are presented that examinee must hold in memory or discriminate
30	Oral Language Evaluation	EMC Corporation St. Paul, MN	speaking; listening; interaction	elementary; Spanish, English	individually administered; student's discussion of supplied stimuli is tape recorded and transcribed

TABLE 1 (cont.)

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
31	Profile of Nonverbal Sensitivity	R. Rosenthal, J.A. Hall, M.R. DiMateo, P.L. Rogers, and D. Archer, *Sensitivity to Nonverbal Communication* Baltimore: John Hopkins University Press, 1979	nonverbal decoding	grades 3–6; high school	group administered; students view videotape or film; multiple-choice response format
32	PRI Reading Systems, Oral Language Skill Clusters	McGraw-Hill, New York, NY 10036	listening; nonverbal decoding	grades k–3	group administered; multiple-choice format
33	SRA Achievement Series	Science Research Associates, Inc., 155 North Wacker Dr., Chicago, IL 60606	listening; auditory discrimination	grades k–3	group administered; paper and pencil, multiple-choice format
34	Sequential Tests of Educational Progress; Listening	Addison-Wesley, Reading, MA 01867	listening	grades 3–12	group administered; multiple-choice format
35	Stanford Achievement Tests; Listening Comprehension	Harcourt Brace Jovanovich, New York, NY 10017	listening	grades 1–6	group administered; multiple-choice, paper and pencil format
36	Stanford Early School Achievement Test	Harcourt Brace Jovanich, New York, NY 10017	listening	grades k–1	group administered; multiple-choice, paper and pencil format

TABLE 1 (cont.)

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
37	Situational Language Tasks	E. E. Conrad, R. K. Rentfrow, K. Meredith, and J. M. Fillerup, *Use of Situational Language Tasks in an Intra-TEEM and TEEM versus Comparison Evluation*. Tucson, AZ: University of Arizona College of Education, 1976.	speaking; listening; interaction	grades 1–3	includes whole-class discussion, and structured and unstructured small group discussion; talk is recorded and transcribed
38	Speech in the Classroom: Assessment Instruments of Speaking Skills	Bureau of Curriculum Services, Pennsylvania Department of Education, 333 Market Street, Harrisburg, PA 17126	speaking; speaking experience; attitudes	grades 1–12	assessment of speaking skills individually administered, others group administered; paper and pencil, multiple-choice format
39	Test of Adolescent Language	PRO-ED, 333 Perry Brooks Building, Austin, TX 78701	speaking; listening	ages 11–18	speaking tests individually administered; listening tests group administered; paper and pencil, multiple-choice format
40	Test of Listening Accuracy in Children	Communication Research Association, P.O. Box 11012, Salt Lake City, UT 84111	listening; nonverbal decoding	grades k–6	group administered; multiple-choice format
41	Torrance Tests of Creative Thinking: Verbal Test	Scholastic Testing Service, Inc., 480 Myer Rd. Bensenville, IL. 60106	creative thinking; speaking	grades k–3	individually administered
42	Utah Test of Language Development (Direct Test Version)	Communication Research Associates, Inc., Box 10012, Salt Lake City, UT 84111	speaking; listening; general language ability	ages 2–14	individually administered

TABLE 1 (cont.)

Instrument Number	Title	Source	Skills Tested	Target Populations	Mode of Administration
43	Vermont Basic Competency Program Speaking and Listening Assessments	Vermont Department of Education; Montpelier, VT 05602	speaking; listening; interaction	grades k–12	variety of simulation tasks and observations conducted in classrooms
44	Wallner Test of Listening Comprehension	N. K. Wallner, "The Development of a Listening Comprehension Test for Kindergarten and Beginning First Grade." *Educational and Psychological Measurement*, 1974, *34*, 391–396	listening	grades k–1	recorded instructions, listening passages, and responses; multiple-choice format; administered to small groups
45	Westside High School Minimum Competency Test	Westside Community Schools, Omaha, NE	speaking	grade 10	students present individual talks to group

Only four of the items are evaluative, and only one makes reference to the overall adequacy of the measure as a tool for assessing communication competence. This approach was adopted for several reasons. First, it is not possible to recommend or condemn instruments without knowledge of the specific purposes for which they are being used. An instrument that is useful for evaluating program or teacher effectiveness may not be adequate for placing students in individualized instruction (Rubin, 1981). Second, in the absence of a consensually acceptable model of competent communication, it is difficult to evaluate instruments' content and construct validity. Objectives and competency lists adopted in one jurisdiction may diverge widely from those that guide test selection in another district. For example, some districts emphasize formal, mechanical aspects of vocal delivery (e.g., Gary Community School Corporation, 1977–1978), while others focus on functional aspects of executing communication tasks (e.g., Vermont Department of Education, 1977). At least for the present, selection of evaluation criteria and instruments should be conducted at local levels in accordance with enlightened community standards. Finally, it is anticipated that the primary users of these reviews will not be speech communication scholars, but rather evaluation specialists and school administrators. The instrument review form reflects the general concerns of this target audience with respect to psychometric adequacy and administrative feasibility.

The section of the instrument review form on validity is concerned with the extent to which the instrument actually measures the skills or knowledge it intends to measure. *Validity* may be determined in many ways and the presence of multiple validity studies using different methods and different target populations strengthens the case that the instrument actually measures what it purports to measure.

Predictive validity deals with the ability of the instrument to predict performance on another measure that is known to be valid and that is theoretically related to the instrument in question. For example, a test of communication competence might be assumed to predict success in jobs that rely heavily on oral communication.

Concurrent validity is similar to predictive validity except that it focuses on the relationship between individuals' performance on the instrument in question and on other instruments that measure the same thing. If a group of students were administered a speaking performance test and were also rated by their speech teacher, then the correlation on these two measures would be a test of concurrent validity.

Content validity indicates the degree to which the content of an instrument represents the domain of knowledge and skills it intends to measure. Content validity is usually determined through expert judgment. One common method is where experts are given a description of the test objectives and then asked to categorize each item by these objectives. Content validity is measured by the degree of agreement among judges in the category assignments.

Building a case for *construct validity* takes many forms. Construct validity includes any experiment that sheds light on the nature of the phenomenon that the instrument is trying to measure. Factor analysis of the items in an instrument is sometimes conducted in order to explore the underlying relationships among the items. Theoretical models about the phenomenon are used to formulate and test hypotheses about how the instrument should operate. For example, several listening comprehension tests could be administered to students along with reading and intelligence tests. If listening is a unique skill, then the listening tests should be more highly correlated with one another than with the tests of the other abilities. A third common method for examining construct validity is the known groups method. Here the instrument is administered to two populations that are known to possess and not possess the knowledge or skill being measured. The degree to which the instrument separates the population into the appropriate subgroups is a measure of construct validity.

The section of the review form on *reliability* reports the measurement accuracy of the instrument. There are various methods for determining reliability. *Test-retest reliability* measures the stability of an instrument over time. Assuming that the respondents have not been exposed to instruction and have not undergone a major growth in the knowledge or skills being measured, they should receive approximately the same score on an instrument at two points in time. This is a measure of test-retest reliability.

In some cases instruments are designed to have alternate forms that are equivalent in content and difficulty. The correlation between individuals' scores on the different forms is a test of *alternate forms reliability*.

Taking the concept of alternate forms reliability a step further, it is possible to think of an instrument as a random set of items, each of which is a "test" of some part of the content domain. The degree to which the respondents' performance on one item is related to their performance on other items is a measure of *internal consistency reliability*.

Tests of performance are markedly different from paper and pencil tests. For these tests, measurement takes place within the person who assigns the rating or score. Here the reliability of the scorer is at issue, not the reliability of the test. *Scoring reliability* is usually assessed by having more than one person rate the same performance. The correlations or percentages of agreement in these ratings is a test of scoring reliability. Usually scorers are evaluated for reliability after training but before they begin rating. However, to insure that scorers remain consistent over time, it is important to check their reliability during the scoring process as well.

As a part of the development of some large scale assessment instruments, *norming* or *criterion setting* studies are conducted and these are also discussed on the review form. For these studies, the instrument is administered to a large number of respondents and the results provide performance benchmarks for future users of the instrument. Norming studies for standardized

achievement test yield charts that transform raw scores into normed scores, most frequently grade equivalence. Standard setting studies are sometimes conducted for tests that measure mastery of specific objectives—criterion referenced tests. Data collected from samples of students is usually compared with data from another source, such as teacher ratings, to determine what test scores represent mastery level. A caution for all norming and criterion setting data is that the characteristics of the original population assessed may be different from the population that the user is assessing.

TYPES OF RESPONSE AND SCORING PROCEDURES

Listening

Multiple-choice formats are the stock-in-trade of standardized testing. Questions are designed so that each has a single correct answer; tests can be graded easily by machine or template without any problems of unreliability in scoring. Item difficulty is readily ascertained and controlled, and test forms can be equated by well-established methods. Two indirect tests of speaking ability attempted to utilize multiple-choice responses (20, 26), but the technique is widely represented among tests of listening proficiency (1, 2, 3, 4, 8, 19, 20, 21, 24, 27, 28, 32, 33, 34, 35, 40, 44). (*Note:* The numbers correspond to the instrument numbers used in Table 1 and instrument reviews in Chapter 2.) Not only are multiple-choice questions used to measure literal comprehension, but also to assess higher order abilities like recognition of speaker's purpose, inference-making, and aspects of critical listening. One of the drawbacks of many multiple-choice listening tests is that students must read printed questions and response alternatives, thus confounding listening ability with reading ability. Some listening tests combat this problem by using tape recorded presentations of questions and response options (21, 25, 26). Others use pictures instead of verbal response options (4, 33, 40).

Another technique employed in some measures of listening skill is behavioral response. In particular, this type of performance measure is used in direction following tasks (1, 11, 16, 34, 42, 43). In general, these tasks approximate normal listening activity, and thus they are more valid than less direct measures. However, in some cases the types of behavioral responses demanded may be quite artificial (e.g., "Place a circle around the second largest square").

Speaking

The most common means for assessing speaking skill are performance rating scales (4, 13, 14, 15, 16, 17, 20, 22, 28, 30, 38, 43, 45). Rubin (1981) discusses a number of factors pertaining to the use of this technique in large scale assessments. Their major disadvantages lie in the potential for unreliable scoring and in the relatively large expenditures of staff time. Some systems seek to avoid the costs of external raters by having classroom teachers evaluate students' typical (22) or elicited (28, 38, 43) speech. This approach would seem to exacerbate the problem of rating error, and beg the question of time allocations.

Alternatives to using performance rating scales in assessing speaking ability are techniques that take particular discourse features as indicators of quality of expression. For example, Loban (1976) and McCaleb (1978) both suggest the use of measures of syntactic complexity for assessing oral proficiency. Some measures of discourse features require speech samples to be transcribed and scored later (4-optional, 30-optional, 37). Others call for on the spot judgment of the presence or absence of specified features, and thus they do not require transcriptions (7, 16, 42). In general, the types of features measured are essentially linguistic such as total number of words, lexical diversity, articulation, and sentence expansion (4, 7, 16, 23, 37, 42). Other evaluation schemes of this type employ a combination of linguistic and whole-text descriptions

(17, 30). Whole-text descriptions include such rubrics as, "Narrative that goes beyond the information given in the pictorial stimulus." Extreme caution is in order, however, in interpreting specific discourse features as indicators of quality. Concurrent validity, wherein such features are shown to predict judgments of overall quality, has rarely been established. Indeed, evidence has accumulated that directly contradicts the use of syntactic complexity in particular, as a measure of quality of expression (Crowhurst, 1979).

When speaking tasks are structured in a way that permits objective measurement of success, it is possible to derive measures of communication effectiveness. For example, it is possible to use "shift of opinion ballots," which ask audience members to indicate their attitudes toward a topic both before and after the delivery of a persuasive speech, to measure the effectiveness of the speaker. Referential communication tasks (Dickson and Patterson, 1979) measure communication effectiveness of a speaker by seeing whether a listener can identify the correct object from an array based on the speaker's description of the object. Effectiveness of small group communication can be evaluated by assigning a unique-solution problem to a group and then recording the accuracy and speed of their solution. These techniques, however, do not elicit uncontaminated measures of individual communication competence because audience characteristics, listener skill, and group composition are factors beyond the control of the speaker and can affect communication success. The effectiveness of some referential communication tasks, however, can be assessed without recourse to measuring listener accuracy. For example, some tasks require the speaker to state the attributes of an object or geometric figure that uniquely describe it. Communication effectiveness is evaluated simply by counting the number of critical features that the speaker identifies (Piche, Rubin, and Turner, 1980).

Conspicuous in its absence from the instruments reviewed is use of interaction coding systems for assessing communication skill. Such devices are based on observations of naturalistic interactions. They include simple sociograms indicating the frequency and direction of communication flow as well as category systems that may classify communicators' messages as constructive or dysfunctional vis-à-vis group maintenance and task functions (Bales, 1953).

CONTENT OF ASSESSMENT INSTRUMENTS

Listening

Listening is not a unitary skill, but it is rather a complex of subskills, each of which is brought into play to greater or lesser degrees depending on the nature of the listening task (Lundsteen, 1979). It is natural, therefore, that tests of listening ability tap a variety of skills. Test users should make sure that the listening test selected conforms to their particular measurement objectives.

Most often, listening tests measure literal comprehension of spoken material (1, 3, 4, 6, 8, 19, 20, 21, 24, 25, 26, 27, 28, 32, 33, 34, 35). It should be noted that comprehension is generally confused with recall or retention, since questions typically follow some extended discourse. Two testing methods alleviate this confusion. The Cloze Test (5) provides verbal context that may lessen reliance on memory. Similarly, tests that deliberately select brief passages and present few questions for each passage (21) may tax memory to a lesser extent.

Many listening tests focus on listening for directions (1, 11, 16, 42, 43), a type of purposeful listening that is readily measured by accuracy of behavioral response (e.g., circling the correct item, drawing the proper path on a map). Other listening skills frequently measured include recognition of speaker's purpose (21, 25, 27), making inferences or interpretations beyond material given (3, 4, 19, 20, 21, 25, 27, 28, 32), and summarizing (25, 43).

A few listening instruments emerge as covering rather unique content. The Listening Comprehension Tests (19) include subtests reflecting ability to interpret paralinguistic cues and also

ability to render social judgments from speech. Several tests not reviewed here also measure sensitivity to paralinguistic cues (Smith-Elliot Listening Test; Davitz and Mattis, 1964). An orientation to functions of listening (for example, to gain information or to evaluate credibility), as opposed to subskills, is displayed by the NAEP Pilot Test (26) as well as the Massachusetts Assessment of Basic Skills (21). The NAEP Pilot Test (26) and the Smith-Elliot Test (Learning Dynamics, Inc.) make some provision to assess comprehension of facial and gestural communication.

Some of the subskills tested by instruments reviewed here are abstracted from any reasonably constructed communication context. These subskills, while critical to communicative listening, are so narrow that they might better be considered receptive language skills. Such receptive language skills include, most prominently, vocabulary (1, 2, 3, 4, 6, 8, 15, 31, 38, 40), syntax (31), and phoneme recognition and discrimination (3, 13, 15, 29, 32, 33, 36, 39, 40). Phonemic discrimination and identification is viewed as essential for reading readiness, but should not be construed as a measure of listening ability.

Outside of the classroom, the bulk of listening activity takes place in the course of interaction. When referential communication tasks assessing skills of both the speakers and the listeners permit free oral interchange (9, 10, 18), these tests approximate interactive communication. Other instruments measure interactive listening skill more indirectly by including conversational speech among their listening passages (21, 25, 26). In general, however, interactive listening is an area calling for vigorous test development efforts.

Speaking

The content of speaking assessment procedures is as varied as that of listening tests. One way in which this content can be categorized is in terms of mode of discourse. At the elementary age level, most tasks are either narrative (4, 43) or descriptive (9, 10, 18, 22). A number of tests designed for non-native speakers also rely on story-telling (15, 17, 30). For older native English speakers, greater variety is evident. The tasks often call for exposition in the form of extended monologues (13, 20, 38, 43, 45). Other modes of discourse include extended persuasive monologues or simulated persuasive conversations (13, 22, 43), telephone conversations (22, 43), introductions (43), and responding to questions in an interview (13, 23, 43).

Speech assessment procedures can be categorized in terms of communication situations as well. In particular, it is useful to examine the types of audiences that are featured in oral performance tasks. Of course, students will be aware of the examiner as an ultimate audience. However, in the majority of instruments reviewed, the examiner is the sole audience to whom students speak. Speakers do not typically communicate in order that their oral proficiency may be evaluated. Indeed, evaluation usually inhibits communication. To the extent that assessment procedures offer no pretense for speaking other than evaluation, these procedures yield inaccurate samples of communication performance.

A single examiner-audience is most natural in interview situations (13, 23, 43). One pitfall of interview situations is that the interviewer may exert overriding influence on students' speech behavior, resulting in considerable unreliability (Mullen, 1978; Hitchman,1966; Bazen, 1978). A single examiner-audience is most anomalous in those situations in which students are called upon to deliver a speech to that individual (13). The problem of unnatural audiences is somewhat relieved by procedures that simulate situations involving realistic speaker/audience relations. These procedures may ask students to simulate an emergency telephone call to a police operator, giving directions to a stranger, or persuading a friend to grant some favor (22, 43). These simulation tasks, however, confuse speech proficiency with role-playing ability.

Group discussion has been accorded great importance as an instructional technique, and British educators have attempted to utilize small group peer interaction as an assessment situation (Barnes, 1980). Of the instruments reviewed here, only two sample naturalistic interaction in

peer groups (37, 43). Dyadic referential communication tasks (10, 12, 18) also approximate natural speaker/audience relations.

Criteria

A final aspect of oral assessment content that requires examination is evaluation criteria. Along what dimensions of quality do the instruments render judgments? Listening tests are primarily concerned with accuracy—accuracy of recall, of following directions, of perceptions about social relationships. Conceivably, tests could be devised that provide information about listening activity, as well. Such instruments would indicate what type of listening (critical, aesthetic, informative) students are engaging in during the course of a stimulus passage, and the degree of concentration or constructive assimilation that characterizes their listening processes.

Speech assessment procedures exhibit a fair amount of consistency in their evaluation criteria. Becker (1962) found that typical speech rating scales reflect only three clusters of judgment despite the fact that they may include a larger number of variously labelled criteria. These clusters are content, delivery, and language. These criteria, with the addition of organization, account for most performance rating scales reviewed in this report (13, 20, 22, 28, 45). Despite this consistency in the nature of criteria, rating schemes differ in the weight accorded each criterion and differ in the manner in which the criteria are defined. In particular, instruments vary in their treatment of language. Some instruments weight language most heavily of the criteria, while others apportion emphasis more equally among dimensions of quality. One rating scale, for example, devotes three of seven items to aspects of language, while the remainder concern delivery factors (13). Procedures that result in single, general impression scores rather than analytic judgments (28, 38) by design provide no guidance in how criteria are to be weighted. The definition of language quality adopted by some instruments stresses conformity to the conventions of standard American English (13, 45). Other instruments, particularly those designed for non-native speakers, convey more detailed information about the types of grammatical structures mastered (15, 16, 37, 39).

Just as some listening tests were characterized as so narrow as to qualify more as tests of receptive language, so are some speaking tests measures of productive language, and not communication. This is certainly true of procedures that ask students to imitate words or sentences in isolation and then apply criteria that evaluate articulation or grammatical interference of a first language (29, 39, 42). It is no less true of procedures that incorporate some communicative context like an interview, and then rate speakers on exclusively linguistic grounds (Mullen, 1978). Merely eliciting language by means of a communicative task does not constitute a test of communication competence (Carroll, 1980). To repeat an earlier caution concerning discourse features, it is risky to assume untested relations between linguistic properties and overall quality of expression.

A few speaking instruments that emphasize language quality criteria reflect the contextual and interactive aspects of communication better than many of the more conventional rating scales. These instruments measure the degree to which language is appropriate or adapted to the demands of the communication task. For example, ratings of a response may depend on the type of question asked (23). Or a test may measure the degree of elaboration, not just simple labelling, that is expected in a response to a narrative task (4, 17, 30). Rating scale items may express communication oriented criteria like "appropriateness" or "intelligibility" rather than formal linguistic properties like sentence structure, standard usage, or correct pronunciation.

ADMINISTRATIVE FEASIBILITY

If measures of speaking and listening proficiency are to be adopted for large scale assessment programs, they must be administratively feasible. They must not consume excessive amounts

of pupil time, must not require unreasonable allocation of personnel for administration and scoring, and must not require highly specialized training for administration, scoring, and interpretation. Unlike other basic skills, however, communication is a complex, interactive behavior. Therefore, tests of communication competence are apt to be more expensive than many other large scale assessment procedures.

Many tests of listening ability are, however, amenable to group administration (1, 2, 3, 5, 6, 8, 19, 21, 24, 25, 26, 27, 32, 35). Even skill at following directions can be assessed in this manner (1, 34, 43). Tape recorded administration instructions and response options (21, 25, 26) not only reduce unreliability and confounding with reading ability, but they also contribute to ease of testing. Some listening measures, on the other hand, allow for a wider range of response modes (11, 15, 16, 43), and these require individual administration.

Assessments of speaking skill conducted as interviews or as extended monologues naturally demand individual administration. Moreover, it is advisable to use multiple raters to insure reliability. This can be accomplished by assigning two staff members for "live" rating, or by tape recording performances for subsequent evaluation by two raters. One instrument attempts to reduce the testing burden by requiring classroom teachers to screen their students based on their typical classroom communication behavior and to refer only those students "in question" for individual assessment (22). However, there is some evidence that suggests that these screen ratings were subject to bias, and they are not reliable.

It is possible to reduce administration costs by using group communication tasks, since a number of students can be evaluated during the course of, say, a twenty minute discussion session (Follard and Robertson, 1976). Similarly, referential communication tasks (9, 10, 18) may also be adapted to simultaneous administration to several dyads. Workers at the University of Wisconsin-Madison Research and Development Center for Individualized Schooling are presently experimenting with a promising application of mini-computers which present stimulus arrays for referential tasks and record accuracy of decoding. The least practical methods of oral examination are those that require subsequent transcription and analysis of speech samples (37). Also, some procedures require raters who are well trained in identifying linguistic structures (7, 16, 42).

TARGET POPULATIONS AND POTENTIAL SOURCES OF TEST BIAS

The instruments reviewed here cover the entire K–12 age range, although the elementary grades receive particular emphasis, especially among commercially developed instruments. Several of the measures include alternate forms that can be administered in English or in Spanish (15, 16, 17, 30). Indeed, it appears that sophisticated advances in communication assessment have emerged from the field of second language testing (Carroll, 1980). Only a single instrument is specifically designated as appropriate for special education populations (11).

Stiggins (1981) discusses a number of sources of bias in communication testing. Instruments vary considerably in their efforts to minimize group bias effects. Some technical manuals document the work of minority group reviewers who examined items in order to eliminate potential bias (26). Other manuals tabulate normative data separately for black and white students (4). It should be noted, however, that differences in central tendency are not, themselves, evidence of test bias. Rather, a test is biased if it over- or under-predicts scores on some independently administered criterion measure (Cleary, 1968). In the absence of criterion measures of communication quality, it is difficult to ascertain test bias. The majority of instruments reviewed here, however, do not address the issue of potential group biases. Indeed, some scoring rubrics assign particular weight to standard English dialect patterns, a procedure that likely places nonstandard dialect speakers at a disadvantage.

LOCALLY DEVELOPED INSTRUMENTS

Developing instruments locally for assessing listening and speaking requires considerable time and effort as well as familiarity with measurement and content concerns. Often it is not feasible to submit locally developed instruments to the same degree of technical review for reliability or validity as commercially developed instruments. However, there are some situations where local development of assessment instruments is desirable. For example, a school district may adopt a set of specific speaking and listening competencies and develop an instructional program directed toward building those competencies. In order to measure its success, the district may find that it is better to develop a test locally that is tailored to its specific competencies than to use existing tests that only measure some of those competencies or that only measure those competencies indirectly. The following brief step-by-step descriptions of the development process provide direction to local agencies that wish to develop their own speaking and listening instruments.

Listening

To determine the listening skills and type of listening tasks that are important, local developers should begin by defining the types of listening skills and tasks that students should be able to perform. In developing this list, developers will find it helpful to review curricular objectives, instructional materials, and teaching practices. They should involve a full range of people concerned with the results of the assessment, for example, teachers, curriculum specialists, administrators, parents, and students. The resulting list may focus on skills that are important to all listening situations, for example, understanding main ideas and details. These skills may be similar to reading comprehension skills. The list may also focus on specific listening tasks that are considered important, for example, listening to directions, listening to a lecture, or listening on the telephone. It is critical that the skills and tasks listed be as specific as possible so that they may be objectively measured.

The next step in developing listening assessment instruments is to assemble stimuli that the students will listen to in the assessment. These stimuli should reflect the listening tasks identified in the first step. Listening material may be drawn from existing sources. Natural listening material such as public service announcements, commercials, or news stories make particularly good material. It is also possible to write material that particularly reflects the tasks identified in the first step. Care should be taken to use material that is relatively short, is interesting to students, and does not reflect a bias toward a particular sex, racial/ethnic, socioeconomic, or geographic group.

The actual production of stimulus material may take two forms. The material may be written in script form so that it may be read aloud by the test administrator; or it may be recorded on audiotape or videotape. The advantage of taped materials is that they guarantee standard administration and allow for variety in stimulus material, such as various voices, conversations, or sound effects.

Several possible types of listening items may be developed. The most typical is multiple-choice items that ask a question about the listening stimulus and provide several possible response options. Another type is short-answer items that ask a question and require the student to write a short response. A third type, used for following direction tasks, presents graphic material, such as a map, and asks the student to complete a certain task like drawing a route onto the map. A variation of this listening item is to describe an object and ask the student to draw the object or to select the appropriate object from a set of pictures. In all cases, item development should follow established standards for item construction that may be found in measurement textbooks.

It is impossible to identify all the possible confusing or problematic aspects of stimulus material or items until they have been field tested with a sample of students who are similar to those who will be assessed. The results of field testing may be used to pick the best stimuli and items. Measurement textbooks provide some simple techniques for reviewing field test data. In addition, field testing provides information about the amount of time it takes most students to complete the items. This information should be used to establish the time limits for the finalized test.

Speaking

Similar to listening instrument development, to determine the speaking skills and types of speaking tasks that are important, local developers must first define the types of speaking skills and tasks students should be able to perform. The steps in this process are the same as they are for listening. The resulting list may focus on specific skills important in all speaking situations, for example, speaking distinctly or speaking in an organized fashion. The list may also focus on specific tasks that are considered important, for example, giving directions, giving a speech, or asking questions. As with listening, it is critical that the speaking skills and tasks listed be as specific as possible so that they may be observed and measured.

Two types of approaches are used in assessing speaking behaviors. First, in an observational approach, the student's behavior may be observed and assessed unobstrusively. Second, in a structured assessment approach, the student may be asked to perform one or more structured speaking tasks, and his or her performance on the tasks is then assessed.

If an observational approach is taken, the developer must decide what speaking behaviors will be observed, for example, asking questions, responding to questions, or speaking in group discussions. In addition, the developer needs to decide how many times each student will be observed and for how long. The observer may be the regular classroom teacher or someone from outside the classroom, such as a teacher from another grade level, a chairperson, or a counselor.

If a structured approach is adopted, the developer must decide on what type of speaking tasks will be used. Also the developer must decide on the setting for the tasks. The student might be asked to perform certain tasks in front of the entire class, in a small group setting, or in a one-on-one situation with the assessor. Again, the assessor may be the classroom teacher or someone from outside the classroom.

Next, a scoring system that describes acceptable and unacceptable levels of performance for the speaking skills or tasks already identified in the first step must be developed. The scoring system may involve a two-point determination: the behavior of interest is either present or absent, the student can be heard or cannot be heard. Alternatively, the scoring system may define a continuum of behaviors that range from lowest to highest: the student is very disorganized while speaking, somewhat disorganized, fairly well organized, or very well organized. However, when a continuum is used, it is necessary to describe each level of the scale in terms of specific behaviors that represent that point in the scale. The resulting scoring system will be used either for observation ratings or structured ratings, as determined previously in the second step.

Once the basic approach is established and the scoring system is developed, it is necessary to train raters in the use of the system. Training should include thorough instruction in the categories in the scoring system, provision of examples of performance that represent the various categories, and opportunities for the raters to practice rating student performance. Raters should have ample opportunity to ask questions about the categories and discuss their practice ratings.

Often training will lead to alterations in the scoring system. It is possible that some initial distinctions made in the scoring system will prove impossible to observe in actual performance. Once the system is finalized and raters are comfortable in their ability to make ratings, raters

should be tested for interrater reliability. They should be given several samples of performance and asked to rate them without discussion. The degree of agreement among the raters is a measure of interrater reliability. Raters should also be trained in test administration procedures—either for observations or for structural assessments.

The final steps in conducting the assessment are data collection and scoring. These activities may happen simultaneously or in stages. Ratings may be made on the spot, or the speaking performance of students may be audiotaped or videotaped and scored at a later time. The advantage of recording performance is that it allows for scoring in a more controlled environment.

In addition to testing interrater reliability at the end of training, the reliability of the ratings also should be checked during the assessment. If ratings are conducted on the spot, it is necessary to have more than one person simultaneously rate students. If ratings are conducted later, it is necessary to have more than one person rate the recordings. Checking the reliability does not have to occur for every rating but should be conducted at random for at least 10 percent of the ratings.

SELECTED RESEARCH AND DEVELOPMENT PRIORITIES

What Should be Tested?

Ideally we should test what we teach. But in most cases adoption of oral communication competencies by state and local educational agencies is a new phenomenon. Moreover, even when such competencies are adopted, the extent and fidelity with which they filter down to classroom practice is unknown. Therefore, a major research priority in assessment is to determine current classroom practices in speech communication. While educators periodically conduct status surveys of classes designated as "Speech" (e.g., Brown, et al., 1979; Rubin, 1980b), the need here is for a more comprehensive survey of all English language arts instruction.

Most probably, we would discover that deliberate teaching of oral communication skills is largely neglected in American public schools. Even in Great Britain, where "oracy" has enjoyed greater emphasis in curriculum documents, little explicit speech communication instruction takes place (Barnes, 1980). Therefore, the content domain of communication assessment probably cannot be defined by what is taught, but by what ought to be taught. As Barnes (1980, p. 125) observes of the British schools, "Any monitoring of oracy during secondary schooling will be proposing a wider range of curricular concerns in oracy than schools presently undertake. . . . Thus, in secondary schools at least, the monitoring of oracy is likely to be leading practice in schools rather than responding to it." We return in a later section on educational utility to the issue of what Wilkinson (1968) terms "washback."

The specification of a content domain for testing, then, exerts impact on instruction. Wiemann and Backlund (1980) describe some of the divergent attempts to define "communication competence." Larson (1978) notes that the definitional problem is the greatest impediment to assessing speech communication. Testers who have accepted the definition of the National Speech Communication Competency Project (Allen and Brown, 1976) in constructing measures of listening and speaking skill were unable to devise suitable items for all of the components specified (McCaleb, 1979; Mead, 1977). Furthermore, all components of communication competence may not be within the proper purview of the public schools. For example, Weimann (1977) includes self-disclosure as among communication competence behaviors, but public schools officials may not believe it is their role to inculcate self-disclosure in their students. Thus, it may not be feasible or advisable to test the entire domain of communication competence (assuming it can be defined in a satisfactory fashion). However, no principles or methods exist for sampling from the content domain.

An especially troublesome issue pertaining to the validity of oral communication tests concerns the role of language knowledge and general verbal ability. Functional communication competence, it is generally agreed, is the ability to use codes (verbal and nonverbal) appropriately in situations (e.g., Larson, et al., 1978). Instruments that measure knowledge of the conventions of standard English are not tests of communication competence, although procedures calling for use of standard English in particular contexts may be appropriate (Rubin, 1980a). Some tests of oral expression utilize communication contexts like story-telling or interviews, but assign scores by emphasizing isolated language skills such as articulation, standard grammar (14), or vocabulary and sentence expansion (34). Some commercially available tests of listening comprehension appear to be little more than measures of general verbal ability (Kelly, 1965).

In summary, we propose the following research and development priorities relating to what should be tested in measures of speaking and listening proficiency:

- Conduct comprehensive surveys of classroom practices in oral language arts instruction.
- Define the content domain of communication competence.
- Delineate components in the content domain that are not appropriate for public school instruction.
- Devise principles for sampling from the content domain.
- Develop measures that distinguish between communication competence and general verbal ability.

How Can Criterion Reference Validity Be Determined?

In general, existing instruments for testing communication competence have not been subjected to studies of concurrent or predictive validity. One reason for this may be the rapidity with which state and local education agencies have needed to set up assessment programs. The lack of accepted criteria against which tests may be validated constitutes another reason. Ability test scores are one source of information about concurrent validity, but they are not satisfactory as the only criteria. Holistic teacher ratings of general communication skills such as those envisioned as the first phase in the Massachusetts Assessment of Speaking Skills (22) might prove suitable for this purpose. However, initial data from this project indicated that holistic ratings might be subject to bias and unreliability. Sociometric analyses using peer interaction data could also serve as criteria for concurrent validity. Criteria for studies of predictive validity could include teacher or job ratings at some later point in time.

Establishing criterion referenced validity seems particularly crucial in assessment tasks that are obviously contrived solely for the purpose of evaluation. Several assessment procedures require students to communicate in role-playing situations (Massachusetts Listening and Speaking Assessment, 1980; Rubin, 1980a). While such procedures permit evaluation of "life role" communication skills, the relationship between role-playing performance and natural communication performance is unknown. Other procedures require interviews or conversations with an assessor (e.g., 14). However, Barnes (1980) notes that in British Certificate of Secondary Education examinations, students display different communication behaviors in peer groups than in private interviews.

In summary, we propose the following research and development priorities relating to criterion referenced validity in measures of speaking and listening proficiency:

- Establish criterion measures for measuring concurrent and predictive validity.
- Explore naturalistic criterion measures for these purposes.
- Investigate criterion referenced validity of contrived communication tasks.

Are Measures Reliable?

Researchers in written composition have recognized for some time the multitudinous sources of inconsistency in writing evaluation (Braddock, Lloyd-Jones, and Shoer, 1963), and refinements in scoring procedures have continued to be a major focus of research and development in that field (e.g., Cooper and Odell, 1977). The field of speech communication, in contrast, seems to have pursued investigations of test reliability less vigorously in the past fifteen years (Rubin, 1981).

Some attention has been given to internal consistency or dimensionality in studies of speech rating scales (Becker, 1962), and it appears to be common practice for commercial tests of listening skill to report this aspect of reliability. But other related issues have not been addressed. For example, choice of topic in assessments of writing skill is a significant factor in students' scores (Rosen, 1969). Yet several locally developed measures of speaking proficiency offer students a choice of topics (e.g., 14, Rubin, 1980a) with no apparent evidence of equivalence between topics. Recently developed tests of listening ability utilize tape recorded stimuli to avoid variation in administration (e.g., Mead, 1977), and Wilkinson (1968) observes that interviewer-assessor idiosyncracies can alter performance in speaking assessments. Several writers have commented that single samples of speech are not reliable indicators of communication competence, and that several samples ranging over a variety of speech functions and situations should be taken for a fair assessment (Barnes, 1980; Hitchman, 1966). At present, though, we lack the sort of precise information concerning the requisite size of a reliable sample that researchers in written syntactic complexity have obtained (e.g., Crowhurst, 1977). Virtually no information is available concerning test-retest reliabilities of speaking and listening assessment instruments. Although individual test developers have no doubt done considerable work in establishing training procedures to engender interrater consistency, these procedures have not been shared in the literature like the corresponding rater training program in written communication (e.g., Diederich, 1974).

In summary, we propose the following research and development priorities relating to reliability in measures of speaking and listening proficiency:

- Determine equivalence of varying topics and communication tasks.
- Determine impact of interactive assessor or test administrator on student performances.
- Determine size and diversity of speech sample required for reliable indication of competence.
- Ascertain test-retest reliabilities of existing instruments.
- Refine and publish methods of enhancing interrater reliability.

What Measurement Techniques Are Presently Available?

Clearly the committee wishes to encourage the development of new measurement techniques or the adaptation of research methodologies for purposes of evaluation. However, it is worthwhile examining some of the strengths and weaknesses of those already available. The most optimal assessment procedures are those that are least intrusive. Naturalistic observation, and even classroom teacher ratings, however, introduce problems of rater bias and problems of consistency in tasks and interactants.

Indirect tests of speaking ability would alleviate many sources of inconsistency. In one notable effort (20), it proved quite difficult to construct suitable items. Moreover, indirect tests may be contaminated by extraneous factors like reading ability and "test-wiseness." Also, such indirect tests are likely to exert deleterious "washback" effects on speech communication instructional practices, and this can lead to a focus on rote knowledge rather than internalized skill (Rubin, 1980a).

Direct tests of listening ability present fewer problems involving consistency, particularly when test stimuli are tape recorded. Even so, measures of listening in conversation are elusive. Interactive listening—where the listener is an equal conversational partner who responds and is ever ready to switch into the role of speaker—probably calls on different skills than procedures in which test-takers listen to a tape recorded conversation, and yet different skills from procedures in which test-takers demonstrate their understanding of oral commands. It is possible that referential communication accuracy tasks (Dickson and Patterson, 1979) could be adapted as effective procedures for assessing interactive listening skills.

Some scoring procedures adopt objective metrics that are taken as indicators of communication quality. Frequently these objective metrics are linguistic variables. McCaleb (1979), for instance, suggests the use of T-unit length, an index of syntactic complexity, as one of several measures of speaking proficiency. Other writers, however, have pointed out that syntactic complexity varies with each communication task, and it is not directly related to quality of expression (Crowhurst, 1979). In addition to linguistic variables, other objective metrics are measures of various features of message content, such as narrative elements not explicitly depicted in a stimulus picture. Little is known of the criterion referenced validity of such objective message variables.

In tests of speaking ability, use of rating scales predominate. Typically, rating scales are applied to either extended talks or interview situations. In British Certificate of Secondary Education examinations, oral reading and conversation are the most common speaking situations to which rating scales are applied (Hitchman, 1968). It would be useful to adapt the use of rating scales to less intrusive, more interactive communication situations like small-group discussions (Barnes, 1980).

In summary, we propose the following research and development priorities relating to presently available measurement techniques:

- Enhance reliability of naturalistic observation procedures.
- Develop measures of listening in interactive situations.
- Establish criterion referenced validity of objective linguistic and message content features.
- Extend performance rating scales to less intrusive, more interactive communication situations like small-group discussion.

Are Instruments Susceptible to Group Biases?

Consistent group differences in test scores are not, in and of themselves, evidence of test bias. Rather, test bias can only be ascertained by determining if an instrument over- or under-predicts a particular group's performance on some criterion measure. As discussed in a previous section, however, we presently lack any universally accepted standards for criterion referenced validity of communication assessment procedures.

Nevertheless, culture bound evaluation materials will likely favor one cultural group over another. Such materials may include culture bound communication contexts (e.g., role-playing a business executive), evaluation criteria (e.g., standard English pronunciation and syntax), and test stimuli (e.g., "Point to the grandfather clock").

A less obvious source of potential bias against particular cultural groups is the very notion of an oral communication assessment. Gay and Abrahams (1973) claim that black youngsters generally construe the requirements of direct questioning by adults differently than do white middle-class children. Similarly, Philips (1970) describes socialization patterns among Native American Indians that render an oral communication assessment as an anomalous communication contest. In addition to biases against particular cultural groups, it is possible that communication assessment procedures may treat particular individuals differentially. Certainly

individuals with organic speech defects will not be subjected to the same testing procedures as others.

However, certain personality traits may likewise cause communication disorders. Most well-known among these is communication apprehension (McCroskey, 1977). Will special provisions be made for communication apprehension, or, if not, will the public schools be committed to "remediating" this condition as a part of their responsibility to prepare students for communication competency assessment?

In summary, we propose the following research and development priorities relating to test bias in procedures for evaluating speaking and listening proficiency:

- Develop criterion measures with which test bias may be determined.
- Identify culture-bound communication contexts, evaluation criteria, and stimulus materials.
- Determine the degree to which oral communication assessment is inherently biased against particular cultural groups.
- Clarify the status of personality traits vis-à-vis test bias.

Should Communication Competence Be Assessed?

One reason to assess communication competence is not to certify proficiency among individual students, but to evaluate oral communication instruction (McGlone, 1973). Of course, if our supposition of negligible communication instruction is accurate, then this motive is obviated. Another reason for assessment, however, is to encourage and guide the innovation of speech communication instruction. Testing tends to legitimize a teaching field, and test specifications may "washback" (Wilkinson, 1968) to instructional practice. Thus, important questions pertain to effects of communication assessment on teachers' and administrators' attitudes toward the legitimacy of speech communication, effects on curricular innovation, and effects on classroom practices. These questions concern the utility of measurement efforts.

One negative effect of any testing program is the deterioration of student attitudes. Partly this is a function of the ends to which test results are put. Given the generally dubious psychometric adequacy of most present speaking and listening instruments, it would seem rash to use them for decisions of great consequence. In any event, it is worthwhile investigating whether any potential benefits of evaluating communication skills are off-set by negative attitudinal outcomes.

Because large scale assessment of speaking and listening skills is not widespread, little information is available concerning its effects on institutional allocations. What are the costs of oral communication testing in terms of instructional hours lost, personnel hours expended, and dollars spent? It is indeed likely that many administrators do not encourage large scale direct measurement of speech communication competency because they fear it will be too costly. We lack cost-effectiveness studies such as those that have been conducted in conjunction with direct evaluation of writing ability (e.g., Hudson and Veal, 1979). Again, such cost estimates must be weighed against presently unquantifiable utility.

In summary, we propose the following research and development priorities relating to the utility and advisability of procedures for assessing speaking and listening proficiency:

- Ascertain whether measures are sensitive for purposes of assessing instruction impact.
- Determine effects of assessment programs on teachers' and administrators' attitudes toward communication education.
- Determine the curricular and instructional results of assessment programs.
- Identify the ends toward which test results are put.
- Ascertain effects of evaluation of students' attitudes toward communication.

II. Reviews of Oral Communication Assessment Instruments

This chapter provides individual reviews of forty-five testing instruments that assess oral communication skills. In some cases, the primary purpose of an instrument is to assess a particular oral communication skill. In other cases, an instrument assesses some facets of oral communication in conjunction with other skills.

1. Brown-Carlson Listening Test

AGE RANGE: Secondary, college, and adult.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Fifty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test assesses: (1) immediate recall, (2) following directions, (3) recognizing transitions, (4) recognizing word meanings, and (5) lecture comprehension. The test administration is oral. Two forms are available (Am and Bm). Each form includes seventy-six multiple-choice items.

NORM/CRITERION DATA: The test was normed on a sample of approximately 8,000 secondary level students and 300 college freshmen. The high school sample was fairly representative of the national population with respect to age and ability level.

VALIDITY

Predictive: The test correlated .21 and .28 with high school rank and .41 with honor point ratio.

Concurrent: Correlations with tests of mental ability ranged from .69 to .78 among high school students and from .22 to .55 among college students. Correlations with reading tests ranged from .47 to .66 among high school students and from .31 to .38 among college students.

Content and Item Selection: Test content was based on professional criteria, research, and expert judgment. A large sample of items was originally created and the best items were selected based on the results of a series of field tests.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Alternate forms reliability yielded a median estimate of .78.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Split-half correlations ranged from .84 to .90.

EVALUATIVE REACTIONS

Practicality: The test is simple to administer, it provides useful subscales, and seems practical for both classroom and research use.

Validity For Specific Purposes and Populations: The test has been used with a large number of different cultural, ethnic, educational, and social groups. Its subscales provide specific information about component skills.

Reliability: Evidence indicates test reliability is good.

Overall Adequacy: The test may be tapping more into general intelligence than it is into listening. It is possible to view the two constructs as independent. This test does not do that.

MATERIALS REVIEWED: Brown, J. I., and Carlson, G. R. *Brown-Carlson Listening Comprehension Test*. New York: Harcourt Brace & World, 1955.

OTHER REFERENCES: None.

REVIEWER: John Daly

2. California Achievement Test; Listening for Information; Level 10

AGE RANGE: Kindergarten.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Twenty-five minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures (1) school vocabulary, (2) terms related to space, direction, and location, and (3) the relationship between facts and concepts. It contains sixteen multiple-choice items. The examiner reads a short story to the students. The students are asked to pick the picture out of three choices that answers a question about the story.

NORM/CRITERION DATA: The test was normed on a national sample of public and Catholic school students. About 200,000 students were involved overall. Norming was conducted in both fall and spring.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is group administered. Easy to follow instructions are provided. Numerous scoring procedures are available.

Validity For Specific Purposes and Populations: Inadequate information to judge. Additional information may be available or forthcoming.

Reliability: Inadequate information to judge. Reliability information is published in a technical manual but was not available to the reviewer.

Overall Adequacy: Listening is a small priority in this achievement series, and is only assessed in the first level as a prereading skill.

MATERIALS REVIEWED: *California Achievement Tests*. Monterey, CA: CTB/McGraw-Hill, 1977.

OTHER REFERENCES: None.

REVIEWER: J. C. McCrosky

3. CIRCUS Listening Test

AGE RANGE: Kindergarten–grade 3.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Hand scored edition $11.25; machine scored edition $18.50. The basic package includes thirty-five booklets, teacher's edition, and user's guide.

TIME REQUIRED FOR ADMINISTRATION: Forty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures ability to listen to a story, understand and interpret events in it, remember sequence of events, and understand vocabulary. All items are drawn from a story read by the teacher about a circus. The story is presented in parts and the child marks one of four pictures in response to a question. Two forms of the listening test are available (C and D).

NORM/CRITERION DATA: A sample of over 15,000 children was used for norming Form C. Over 14,000 children were assessed for norming Form D. Data used for norming were weighted according to variables by which the sample was stratified: region, size of community, socioeconomic status, and proportion of minority population.

VALIDITY

Predictive: Samples of children who took one level of the test in the fall were administered a higher level test the following spring. The correlation was .75.

Concurrent: Teacher ratings were collected through the *Child Competency and Learning Inventory* as an independent measure of the same abilities measured by the test. Correlations between teacher ratings and the listening test were .47 on Form C and .43 on Form D.

Content and Item Selection: Item selection was based on the judgment of early childhood experts with a view toward assessing domains which are of interest to teachers and which can be effected by curricula. Each form went through two pretest examinations. Final items were selected based on rigorous evaluation.

Construct and Other Empirical Studies: The authors point to the relationship among various forms of the test as evidence of construct validity. Correlation between Forms C and D was .75.

RELIABILITY

Alternate Forms: The correlation between Forms C and D was .75.

Test-Retest: Children were administered separate forms of the test in fall and spring. The correlation was .75.

Scoring: Not applicable.

Internal Consistency: The average inter-item correlations for Form C were .85, .80, .79 and for Form D were .79, .78, .81.

EVALUATIVE REACTIONS

Practicality: Group administration and optional machine scoring makes this test practical for large scale assessment. Instructions are clear and the test can be easily administered by teachers.

Validity For Specific Purposes and Populations: Evidence indicates that test validity is good. However, this test does not provide an opportunity for children to listen and respond in a conversational way.

Reliability: Evidence indicates that test reliability is very good.

Overall Adequacy: This is a well designed test with a rigorous research base, for assessing general school readiness. It is *not* a test of speech communication ability, but a paper and pencil test. Listening measured in this manner correlates with reading ability. The relationship to the ability to talk with or inform others is unknown.

MATERIALS REVIEWED: Educational Testing Service, *CIRCUS Listening Test*. Reading, MA: Addison-Wesley Publishing Co., 1979.

OTHER REFERENCES: None.

REVIEWER: Janice Patterson

4. CIRCUS Say and Tell

AGE RANGE: Preschool–grade 3.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: $5.50 for ten booklets.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures productive language. There are four levels. Level A is appropriate for preschool and kindergarten. Levels B through D are identical and appropriate for kindergarten through grade 3. Each level has three parts. The test is individually administered. For Part I the child describes two objects: one in a structured response situation and one in a free response situation. In Part II, the child is shown pictures and is asked to generate responses that require the correct use of plurals, verb tenses, prepositions, subject-verb agreement, comparatives, possessives, and conjunctions (for levels B, C, and D). Responses to Parts I and II are scored as correct, partially correct, or incorrect based on protocols. For Part III the child is shown a picture and asked to describe it. Responses to Part III are scored in terms of number of words, number of different works and presence of several qualitative criteria, such as "naming at least four objects or characters."

NORM/CRITERION DATA: Level A was normed with a sample of 227 preschoolers and 541 kindergarten students. Level B was normed with a sample of 805 students mostly ages 6 and 7. Additional norm data may be available.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: Average inter-item correlations range from .49 to.90 for the various parts of Form A. Additional information on other forms may be available.

EVALUATIVE REACTIONS

Practicality: The test must be individually administered. However, clear directions are provided. Scoring is somewhat cumberson.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Evidence indicates test reliability is adequate.

Overall Adequacy: The test provides an adequate sample of children's productive language.

MATERIALS REVIEWED: Educational Testing Service, *CIRCUS Say and Tell.* Reading, MA: Addison-Wesley, 1979.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

5. Cloze Listening Test

AGE RANGE: Secondary.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: Tapes cost $25 per set. Test forms cost $1.50 per form.

TIME REQUIRED FOR ADMINISTRATION: Twenty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures: (1) recall of specific information, (2) ability to grasp the thought of the passage as a whole, (3) ability to apply the limited number of contrastive units which identify the word patterns, and (4) grammatical structures of spoken American English. A short fictional episode is read aloud (about ten minutes in length). Then several excerpts are read aloud by the same narrator. Within each excerpt several words (nouns and main verbs) are replaced by a chime. Students write the missing words on their response sheets. About 40 percent of the selection is included in each excerpt. The test includes two forms (Lisbon and Waco).

NORM/CRITERION DATA: The test was normed on 636 students in ten runs.

VALIDITY

Predictive: No information provided.

Concurrent: With 107 students, the Lisbon form correlated .71 with the Brown-Carlson Listening test, Form Am, Part E, Lecture Comprehension.

Content and Item Selection: The test was reviewed by ten curriculum specialists who judged that the test measures the content it intends to measure.

Construct and Other Empirical Studies: With forty-six subjects, the Lisbon form correlated .79 with the Terman-McNemar Test of Mental Ability, Form C.

RELIABILITY

Alternate Forms: With eighty-three students, the two forms correlated .92. With 130 students, the two forms correlated .87.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: For the ten norming runs, average inter-item correlations ranged from .83 to .96.

EVALUATIVE REACTIONS

Practicality: The test is easy to administer and score.

Validity For Specific Purposes and Populations: Evidence indicates test validity is adequate.

Reliability: Evidence indicates test reliability is very good.

Overall Adequacy: The test only measures exact recall. It does not measure higher level listening comprehension skills.

MATERIALS REVIEWED: Bowdidge, J. S. *Cloze Listening Test*. Springfield, MO: Drury College, 1967. (ERIC Document Reproduction Service No. ED 091 761)

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

6. Comprehensive Tests of Basic Skills, Tests 2, 3, and 4

AGE RANGE: Early elementary.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- X Visual Decoding
- X Subskill or Attitude: auditory discrimination

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Test 2, twenty-one minutes; Test 3, nineteen minutes; Test 4, twenty-one minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Test 2, Sound Recognition, measures sound recognition in three ways: (1) by saying two words (e.g., lake . . . lake or cap . . . cup) and asking the student if the words are the same or different, (2) by asking which word (stated and shown in pictures) begins with the same sound as the stated word, or (3) by asking which word (stated and shown in picures) rhymes with the stated word. In Test 3, Reading Vocabulary, the student is given an oral definition and has to match it with a picture or a word. In Test 4, Reading—Oral Comprehension, the student hears a story and has to answer a question about it by picking the correct picture.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is easy to administer and uses machine scoreable answer booklets.

Validity For Specific Purposes and Populations: Validity studies are in process but at this time information is inadequate to judge.

Reliability: Reliability studies are in process but at this time information is inadequate to judge.

Overall Adequacy: The test is designed as a reading readiness test. It is less useful as an oral communication test.

MATERIALS REVIEWED: *Comprehensive Test of Basic Skills*. Monterey, CA: CTB/ McGraw Hill.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

7. Communicative Evaluation Chart from Infancy to Five Years

AGE RANGE: Infant–kindergarten.

SKILLS TESTED:
- X Speaking
- X Listening
- X Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- X Subskill or Attitude: physical development and visual-motor-perceptual skills

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is divided into two parts: language use and physical development. The purpose of the test is to determine quickly if a child should be referred to a specialist for further testing, therapy, or education. Items that assess language focus on (1) coordination of the speech musculature, (2) development of hearing-acuity and auditory perception, (3) acquisition of vowels and consonants, and (4) growth of receptive and expressive language. The test administrator indicates + if the skill is present, − if not present, ± if it fluctuates. Numerous minus or fluctuation markings indicate further evaluation may be necessary.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: Some items were compiled from sources such as Gesell, Benet, Caltell, and others. Other items were included because they were believed to be diagnostically significant in working with young children.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test requires no training to administer. It is inappropriate for large scale assessment because it calls for individual evaluation.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: The test is judged to be poor because there is no explanation of item development, validity, reliability, norming, and score interpretation.

MATERIALS REVIEWED: Anderson, R. M.; Miles, M.; and Matheny, P. A. *Communicative Evaluation Chart from Infancy to Five Years*. Cambridge, MA: Educators Publishing Services, Inc., 1963.

OTHER REFERENCES: None.

REVIEWER: Janice Patterson

8. Durrell Listening-Reading Series: Primary, Intermediate, Advanced Levels

AGE RANGE: Primary, grades 1–3; Intermediate, grades 4–6; Advanced, grades 7–9.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: $15.75 to $19.65 for thirty-five copies.

TIME REQUIRED FOR ADMINISTRATION: Primary, seventy minutes; Intermediate, eighty-five minutes; Advanced, eighty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures comprehension in both the listening and reading modes. The test has three levels: primary, intermediate, and advanced. Each level is available in two forms (DE and EF). Each level has four parts: (1) Listening vocabulary, (2) listening comprehension of sentences (primary) or of paragraphs (other levels), (3) reading vocabulary, and (4) reading comprehension of sentences (primary) or of paragraphs (other levels). The listening and reading tests are parallel in content and difficulty to allow for comparisons. The listening tests do not require reading or writing ability for responses.

NORM/CRITERION DATA: Each level was standardized and normed on a population of students with normal intelligence and population from average socioeconomic and educational backgrounds. The sample consisted of 22,247 students.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: Vocabulary words were selected to represent categories in Roget's *Thesarus* and assigned to levels based on word lists and field test results. Standard item statistics were used in selecting items.

Construct and Other Empirical Studies: The listening subtests correlated .47 and .52 with the Metropolitan Reading Readiness test in grade 1, from .15 to .65 with the word knowledge and reading subtests of the Metropolitan Achievement tests in grades 2 through 6, and from .48 to .76 with the Iowa Test of Basic Skills in grades 3 through 6. The correlations between the reading subtests and the other reading tests listed above were higher than those between the listening subtests and the other reading tests.

RELIABILITY

Alternate Forms: Listening and reading tests were equated and then reversed to create the alternate form. However, no specific alternate forms studies were cited.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Split-half correlations ranged from .92 to .97 for total listening in grades 1 through 8.

EVALUATIVE REACTIONS

Practicality: The test is group administered. The manual gives specific instructions.

Validity For Specific Purposes and Populations: Evidence indicates that listening is somewhat different from reading which suggests that the listening score might not be a good measure of reading potential—a use that is promoted by the authors.

Reliability: Evidence indicates test reliability is very good.

Overall Adequacy: The test is limited in its coverage of listening skills. The test emphasizes tasks that are similar in listening and reading and does not address tasks that are more typical of just the listening mode.

MATERIALS REVIEWED: Durrell, D. D., and Hayes, M. T. *Durrell Listening-Reading Series*. New York: Harcourt, Brace, Jovanovich, 1970.

OTHER REFERENCES: None.

REVIEWER: J. C. McCrosky

9. Dyadic Task-Oriented Communication

AGE RANGE: Elementary.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is a proficiency test for non-native English speakers. Students are paired. One student is given a task and the other student must respond. Individuals are tested with several partners. Responses are tape recorded. Tasks include requests, manipulative instructions, and descriptions. In general, criteria for evaluation are (1) time and accuracy and (2) comparison with native speakers. Tasks are described and graded by difficulty. Some specific scoring criteria are provided.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: No information provided.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is practical for use in a classroom setting.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: This type of test could be used with native speakers also. Tasks would have to be harder and criteria made more difficult.

MATERIALS REVIEWED: None.

OTHER REFERENCES: Findley, C. A. Dyadic Task-oriented Communication: Exercises for Teaching and Testing in the Elementary ESL Class, 1977. (ERIC Document Reproduction Service No. ED 145 629)

REVIEWER: Nancy Mead

10. DYCOMM: Dyadic Communication

AGE RANGE: Not specified.

SKILLS TESTED:
- X Speaking
- X Listening
- X Interaction
- X Visual Encoding
- X Visual Decoding
- ___ Subskill or Attitude

COST: User produces materials and scores.

TIME REQUIRED FOR ADMINISTRATION: Described below.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures communication skills in five areas: (1) word identification, (2) sentence processing, (3) giving-receiving directions (referential or informing skills), (4) interpreting affect, and (5) problem solving. For each task a group of ten or more people work in dyads. After each task, they rotate to a new partner.

The test for word identification skills is a paper and pencil task in which the dyad communicates so that each may correctly identify the target word. Correct responses receive one point, incorrect answers lose a point, and no points are scored for blank items. Each individual has a five second trial as speaker and five seconds as listener before rotating to a new partner. The tester selects words according to population characteristics.

The sentence processing task requires speaker and listener to communicate about sentences and decide if they are similar or different. Each dyad considers twelve items (for a total of twenty seconds) prior to rotating to a new partner. The score is the number of sentences marked correctly.

The giving-receiving directions task calls for the dyad to be seated in a circle facing a partner as they work to identify abstract figures. The drawings are on a score sheet which is vertically divided into two parts. On the left side of the sheet, target pictures are presented which each person in the speaker role will describe. On the right side of the sheet are several sets of five pictures; the pictures in each set are similar. The speaker describes the target picture and the listener responds by naming the number of that picture on his sheet; both speaker and listener record this number on their papers. The members of the dyad alternate roles every thirty seconds. After each dyad has had a turn as both speaker and listener, they rotate to new partners. The dyad scores a point for each correctly identified picture. Pictures used for this activity are to be selected by the DYCOMM user to insure their appropriateness to the population using the activity. Sample drawings are included in the description.

The fourth task, affects, again calls for children to work with dyads. They discuss a work sheet in a way which sends an affective as well as a cognitive message. The task is for the

listener to identify the correct affective tone. Scoring is simply the number of correct responses. Separate scores are tallied for speaker and listener.

The problem solving task requires the dyad to discuss rules about numbers so they can correctly identify numbers on a worksheet. An example of a rule is "The first digit times the third digit must be more than 10." There are four rules for each task, with two of the rules on the speaker's sheet and two on the listener's. The dyad works forty-five seconds on each item and then rotates to new partners. The dyad scores one point for each correct choice.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: The tester is instructed to select specific words, sentences, emotions, tasks, and problems appropriate to the target population.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: This test is inappropriate for large scale administration due to a lack of available materials.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: This test is a poor measure of communication accuracy due to a lack of systematic evaluation. The test provides good instructional techniques because students are encouraged to talk with each other. The activities will appeal to many age levels.

MATERIALS REVIEWED: Byers, B. H. *DYCOMM: Dyadic Communication*. Honolulu, HI: University of Hawaii, 1973.

OTHER REFERENCES: None.

REVIEWER: Janice Patterson

11. The Fullerton Language Test for Adolescents

AGE RANGE: Ages 11–18, normal and speech impaired.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- X Subskill or Attitude: auditory synthesis, morphological competence, homonyms, enumerating members of classes, syllabification, understanding idioms, grammatical judgments

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Forty-five minutes per subject.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Subtest 3 relates to listening and requires the student to receive, retain, interpret, and demonstrate an understanding of oral commands. The entire test is individually administered. The twenty items range along a dimension of increasingly complex syntactic constructions and logical operations. Commands concern manipulations of colored geometric shapes, included in the testing kit. Responses are scored dichotomously.

NORM/CRITERION DATA: The test was normed with 762 subjects aged 11–18 from regular classrooms in California and Oregon. Based on this sample, a "competence level," "instructional level," and "frustration level" is defined in terms of standard deviations from the mean on each subtest.

VALIDITY

Predictive: No information provided.

Concurrent: The scores on the test distinguished between a normal population (N = 489) and a special education population (N = 73).

Content and Item Selection: Content validity was established by theoretical rationale and by comparison with similar instruments.

Construct and Other Empirical Studies: The test discriminates between normal and special education populations. Correlations between oral commands and other subtests ranged from .37 to .55.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: The test-retest correlations exceeded .80 for all subtests.

Scoring: Not applicable.

Internal Consistency: The average inter-item correlations exceeded .70 for all subtests.

EVALUATIVE REACTIONS

Practicality: The oral commands subtest is easily scored, administered, and is not too time consuming, however, it requires individual administration. The entire Fullerton battery requires forty-five minutes.

Validity For Specific Purposes and Populations: The test appears useful for distinguishing between normal and abnormal language development among adolescents. Little justification is given for interpretations of test scores (e.g., competence level).

Reliability: The reliability of the test is quite adequate, assuming trained administrators are consistent in conducting the test.

Overall Adequacy: The test assesses only a limited type of listening ability using contrived and artificial speech stimuli.

MATERIALS REVIEWED: Thorum, A. R. *The Fullerton Language Test for Adolescents (Professional Manual)*. Palo Alto, CA: Consulting Psychologists Press, 1980.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

12. Fundamental Achievement Series, Verbal

AGE RANGE: Adolescents and adults, with limited educational opportunities.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: No information provided.

TIME REQUIRED FOR ADMINISTRATION: Thirty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The verbal test measures a variety of language skills needed for employment, including vocabulary, reading comprehension, listening comprehension study skills, copying, and spelling. The test has two forms (A and B). The entire test is presented orally but only ten items directly measure listening comprehension. These items are literal comprehension questions about three brief announcements.

NORM/CRITERION DATA: The test was normed with groups of between 100 and 200 individuals. Groups included whites and blacks in grades 6, 8, 10, and 12 and numerous industrial and anti-poverty program groups.

VALIDITY

Predictive: No information provided.

Concurrent: A number of small concurrent validity studies using criteria such as supervisors', researchers', and counselors' ratings indicated statistically significant correlations with the verbal test.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: The verbal test correlated with various tests of general mental ability .36 to .94 indicating that the test is operating in the same general area of measurement.

RELIABILITY

Alternate Forms: In one study of thirty-nine anti-poverty program participants, the verbal test forms A and B correlated .74 with two weeks intervening.

Test-Retest: Scores from two administrations correlated .62 to .95 for five groups of industrial and anti-poverty program participants with three months intervening. However, in all studies the means increased over time.

Scoring: Not applicable.

Internal Consistency: Average inter-item correlations for the verbal test ranged from .70 to .96 for various groups.

EVALUATIVE REACTIONS

Practicality: The test is group administered. The entire test is presented on a tape recording, making it very easy to administer.

Validity For Specific Purposes and Populations: The test is in an experimental stage. However, numerous small validity studies indicate that it is a good measure of basic skills and that it works well with disadvantaged populations.

Reliability: Evidence indicates reliability for the total test is good.

Overall Adequacy: The test only measures a small domain of listening skills and this section would not stand on its own as a unique measure.

MATERIALS REVIEWED: *Fundamental Achievement Series*. New York: The Psychological Corporation, 1964.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

13. Gary, Indiana Oral Proficiency Examination

AGE RANGE: Grade 10.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: No materials required; only record keeping costs.

TIME REQUIRED FOR ADMINISTRATION: Variable, probably no more than five minutes per student.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Two weeks prior to testing students select one of four test formats: (1) the interview format, (2) the topic format, (3) the question format, or (4) the prepared format. The interview format may pertain to personal interests, education, or biographical data and includes such questions as, "Do you plan to marry? If so, specify." and "Where was your father born?" The topic format includes five subjects such as, "How can a person help to improve his or her school, community, or country?", one of which students are assigned at the time of testing. In the question format, students may choose to answer one of five queries such as, "Why are 'rules' made in our homes, schools, or country?" For the prepared format, students prepare an original two-minute speech about a subject of their choice, and deliver it "without notes or crutches." Multiple raters use a holistic scale with four-interval items ranging from "severely deficient" to "moderate to high proficiency." The scoring categories are: (1) articulation, (2) pronunciation, (3) verbal utterances (e.g., "you know"), (4) rate, (5) standard word usage, (6) voice quality, and (7) volume.

NORM/CRITERION DATA: No information provided. However, raters were apparently trained by means of samples which exemplify each score/level.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: The criteria directly reflect the program's oral proficiency performance objectives.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided. However, raters are apparently checked frequently for wide discrepancies.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: Testing should progress rapidly, especially since students have opportunity to prepare long in advance of testing. Interview formats allow little opportunity for spontaneous elaboration or interaction which might be time consuming.

Validity For Specific Purposes and Populations: Criteria are well suited to objectives, particularly as they are realized in formal, pressured contexts. Criteria seem to be biased against speakers of nonstandard dialects.

Reliability: The major problem is the unknown equivalence between formats and topics. Some provision seems to have been made for establishing and maintaining rater reliability.

Overall Adequacy: The contexts are highly artificial and nonmotivating, despite choice accorded to students. Criteria reflect a narrow range of competencies restricted to elements of elocution.

MATERIALS REVIEWED: Gary Community School Corporation. *Oral Proficiency Program*. Gary, IN: author, 1977–1978.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

14. Glynn County Speech Proficiency Examination

AGE RANGE: Secondary.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: No commercial materials needed.

TIME REQUIRED FOR ADMINISTRATION: Approximately one hour per twenty students.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Using scripted instructions, the test administrator sets up a situation that simulates a public hearing before a county board of education. Students are provided with an agenda that includes some background information about three selected issues. Students present persuasive speeches individually. Two raters score either videotapes or live performances with discrepancies resolved by a third rater. Rating scales specify four skill level indicators for each of the following dimensions: (1) introduction, (2) purpose, (3) reasons, (4) organization, (5) objections, (6) conclusion, (7) language style, (8) oral expression, and (9) gestures.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: For a sample of thirty ninth graders, scores on this test correlated .70 with a parallel form of the test which involved performance on a job interview task rated by slightly different criteria.

Content and Item Selection: The task and rating criteria conform to locally stated objectives.

Construct and Other Empirical Studies: Ratings of this test correlated .88 with classroom teachers' judgments of students' typical communication competence. There was also a highly significant relationship between passing performance on this instrument and students' placement in ability level tracks by the school system.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Inter-rater reliability was .82 when using videotaped performances and .72 when scoring live performances. When considering passing versus nonpassing scores, 15 percent of the students were cross-classified by two raters and required a third rating for resolution.

Internal Consistency: Average inter-item correlations ranged from .82 to .88.

EVALUATIVE REACTIONS

Practicality: The test is relatively time consuming in both administration and rating. The authors estimate .2 personnel hours per student. Also the test consumes a large amount of "down time" for students.

Validity For Specific Purposes and Populations: While rating criteria appear well tailored to the speaking task, this measure samples only a limited range of speaking competencies. The cultural bias of the test is unknown.

Reliability: Raters must be trained to achieve reliability and a standardized regimen of training would need to be developed. Topic and speaking order effects appear to be insignificant. Test-retest reliability is unknown, but likely a troublesome point.

Overall Adequacy: The test represents strong effort at speech performance assessment. The measure attempts to create a sense of context. However, a single speech sample representing just one communication situation is not representative of general speaking skills.

MATERIALS REVIEWED: Rubin, D. L., and Bazzle, R. E. *Development of an Oral Communication Assessment Program: The Glynn County Speech Proficiency Examination for High School Students*. Brunswick, GA: Glynn County School, 1981.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

15. Language Assessment Scales

AGE RANGE: Grades 1–5.

SKILLS TESTED:
- X Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- X Visual Decoding
- ___ Subskill or Attitude

COST: $56.50.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test includes five subscales assessing linguistic proficiency in either Spanish or English. The phonemic discrimination subscale has thirty items. Subjects determine if two words in which a phoneme or allophone is embedded sound the same or different. The phoneme production subscale has thirty-six items. Subjects imitate words or short sentences in which a sound is embedded. Both are scored right or wrong. One lexical ability subscale, with twenty items, has children identify words for objects presented in pictures. The oral production subscale has children orally retell a story that is cued with pictures. It is scored with a five-point rating scale. Finally, vocabulary is assessed. Procedures for this subscale are unclear in the manual.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: The test predicted language achievement better than cognitive style/development variables. The test accounted for 40 percent of the variance in language achievement scores.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: In a factor analysis the subscales of the test were contributed to the same factor.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: Scores for two administrations with about one week intervening correlated .88 in English and .97 in Spanish.

Scoring: The inter-rater reliability, where used, is very high.

Internal Consistency: The internal consistency reliability is quite high.

EVALUATIVE REACTIONS

Practicality: The test is useful for bilingual programs.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Evidence indicates test reliability is very good.

Overall Adequacy: The test is narrow in what it assesses. It is very adequate for the dimensions it does assess.

MATERIALS REVIEWED: DeAvila, E. A.; Ulibarri, D. M.; Duncan, S. E.; Fleming, J. S.; Costa, M.; Perry J.; and Wainwright, C. Predicting the Academic Success of Language Minority Students from Developmental, Cognitive Style, Linguistic and Teacher Perception Measures, 1979.

OTHER REFERENCES: None.

REVIEWER: John Daly

16. Language Dominance Survey

AGE RANGE: Kindergarten–grade 12.

SKILLS TESTED:
- X Speaking
- X Listeninng
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures students' abilities to: (1) understand and comply with stated commands and (2) communicate with acceptable morphology and syntax. It is designed to identify children who need bilingual education. Consequently, both English and Spanish is used in the test.

NORM/CRITERION DATA: The manual indicates that students scoring less than 50 percent correct should be placed in a bilingual program. Justification for this recommendation is not provided in the report.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: This test is practical for identifying students who should be placed in bilingual classes.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: The test is limited to very basic language skills.

MATERIALS REVIEWED: Burt, M. K., and Dulay, H. C. *Language Dominance Survey*. Berkeley, CA: BCDEL/Lau Multilingual Center, 1974–1975.

OTHER REFERENCES: None.

REVIEWER: John Daly

17. Language Facility Test

AGE RANGE: Ages 3–15, normal populations.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: $20.00.

TIME REQUIRED FOR ADMINISTRATION: Approximately ten minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The child is given three pictures, one at a time, and asked to tell a story about each one. Each response is scored on a 0–9 scale with 0 being no response, 1 being a one word response, 2 being a multiple word response, 3 being a complete sentence, and so forth. The highest response, 9, is an organized, complete story. The student is given several probes if necessary. The score is the sum of the ratings on the three pictures. Three alternate forms of pictures are available—photographs, line drawings, and reproductions of Spanish art masterworks. The test may be given in the student's native language, sign language, or English. Success is based on elaboration of language, not on standard grammar or vocabulary.

NORM/CRITERION DATA: The test was normed on 4000 students ages 3 to 20. Smaller studies were conducted on special groups, including low achievers, mentally retarded, handicapped, and rural Spanish speakers.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: The test correlated slightly with intelligence, reading readiness, achievement, and teacher ratings of scholastic performance.

RELIABILITY

Alternate Forms: The forms correlated .46 to .90 (with intervening instructional activity).

Test-Retest: See above.

Scoring: Inter-rater correlations ranged from .88 to .94 across three pictures.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test requires one-on-one administration and trained scorers. The test is straight forward and does not take long to administer.

Validity For Specific Purposes and Populations: Evidence indicates that test validity is adequate.

Reliability: Evidence indicates that test reliability is adequate.

Overall Adequacy: The test focuses on language and cognitive development. It does not measure functional communication competence.

MATERIALS REVIEWED: *Language Facility Test*. Alexandria, VA: Allington Corporation.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

18. Language Skills Communication Task

AGE RANGE: Kindergarten–grade 2.

SKILLS TESTED:
- X Speaking
- X Listening
- X Interaction
- X Visual Encoding
- X Visual Decoding
- __ Subskill or Attitude

COST: User produces materials and scores.

TIME REQUIRED FOR ADMINISTRATION: No time limit.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is composed of two tasks designed to assess the child's ability to get meaning and ideas from conversation, respond to the language of others, transmit ideas to others, and adapt his or her conversation to achieve effective communication. These are referential or informing tasks. In each instance, two children are seated on opposite sides of a picture board which is the same on both sides. One child is designated speaker and the other listener. The task is for the speaker to tell the listener where to place the pictures scattered loosely in front of the listener on the board. The goal is for the listener's picture to match the speaker's. The children are not permitted to see each other's boards or to make gestures. The listener can ask for more information.

The communication accuracy is scored both individually and as a dyad. The speaker receives one point in each of three areas by supplying specific verbal instructions for: (1) object identification, (2) object placement, and (3) object positioning. The listener's score is determined by (1) selection of correct object, (2) placement of the object, and (3) questioning when insufficient information is given.

Mean scores are calculated for each task to assess dyadic communication. This score is based on the average of the six subscores earned by speaker and listener. The listener's success in placing the object correctly is also a measure of the dyad's communication effectiveness. Children's responses are recorded; these data are used for analysis.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: Children's performance on individual subtests predicted their success in placing the object correctly. Other results indicated that correct object was correlated significantly with grade (.32), math achievement (.43), and reading achievement (.37). Sex and intelligence were not significantly correlated with object placement.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: Twelve first grade children were given the test twice, a week apart. Children were randomly paired and randomly assigned to play the same role for both sessions. The mean percentge of agreement for the task was 89.3 percent, with a range from 78.5 percent to 100 percent.

Scoring: No information provided.

Internal Consistency: Split-half correlation was .73 for the first task and .76 for the second. The scores of twelve first graders were used to determine this measure of internal consistency.

EVALUATIVE REACTIONS

Practicality: The test is inappropriate for large scale administration due to testing procedure, i.e., assessing children in pairs. No special training is necessary for test administration.

Validity For Specific Purposes and Populations: By the author's own admission, more work must be completed to establish validity of the test.

Reliability: Current reliability measures are inadequate.

Overall Adequacy: Given the similarity in the test design to other measures of referential accuracy, this test may provide useful data. However, more rigorous, systematic evaluation is needed before test users can be assured of adequate validity and reliability in the instrument.

MATERIALS REVIEWED: Wang, M. C.; Rose, S.; and Maxwell, J. *The Development of the Language Communication Skills Test*. Pittsburgh: The University of Pittsburgh, Learning Research and Development Center, 1973.

OTHER REFERENCES: Dickson, W. P., and Patterson, J. H. *Communication Education*, 1979.

REVIEWER: Janice Patterson

19. Listening Comprehension Tests

AGE RANGE: Battery A, ages 10–11; Battery B, ages 13–14; Battery C, ages 17–18.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Approximately $75.00.

TIME REQUIRED FOR ADMINISTRATION: Approximately two thirty-minute sessions, with a short break.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test includes five different subtests: (1) content—basic comprehension, (2) contextual constraints—infer missing parts of conversation, (3) phonology—understand differences in meaning brought about by different inflections, etc., (4) register—detect inappropriate uses of language, and (5) relationships—detect kinds of relationships existing between people from language used. Students listen to tape recorded stimuli and answer multiple choice questions about what they heard.

NORM/CRITERION DATA: The test was normed on a sample of 1,152 individuals.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: The test design was based on a theoretical description of listening.

Construct and Other Empirical Studies: The test correlated from .45 to .60 with Simplex Junior Intelligence Scale and AH4 Group Test of General Intelligence. The test correlated from .41 to .75 with Schonell Silent Reading Test, Secondary Reading Test, and Senior Reading Test.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Average inter-item correlations ranged from .78 to .84.

EVALUATIVE REACTIONS

Practicality: The test is very easy to administer in group. Instructions and items are all provided on a tape recording.

Validity For Specific Purposes and Populations: Evidence indicates that test validity is good.

Reliability: Evidence indicates that test reliability is good.

Overall Adequacy: British accents would be difficult for American children. A larger problem is the fact that words and topics would be unfamiliar.

MATERIALS REVIEWED: Wilkinson, A., Stratta, L.; and Dudley, P. *Listening Comprehension Tests*. London: Macmillan Education, Ltd., 1974.

OTHER REFERENCES: Wilkinson, A.; Stratta, L.; and Dudley, P. *The Quality of Listening*. London: Macmillan Education, Ltd., 1974.

REVIEWER: Nancy Mead

20. MACOSA Listening and Speaking Tests

AGE RANGE: Grades 3, 6, 9, and 12.

SKILLS TESTED:
- X Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Listening, sixty to seventy minutes; Speaking (written), forty minutes; and Speaking (oral), thirty-five to sixty-two minutes for groups of six.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: For the listening tests, students listen to tape recordings of various formal and informal listening material and answer multiple-choice questions about what they heard. Questions measure voice production factors, linguistic factors (words and sentences), and organizational factors (literal, interpretative, and critical comprehension).

The speaking tests include both written and oral components. The written tests include multiple-choice questions which measure knowledge of how voice production, nonverbal, and linguistic factors can convey meaning and (at grades 9 and 12) knowledge of speech organization. The oral tests assess articulation (at grade 3 only), spontaneous speech, and spontaneous-prepared speech (at grades 9 and 12). Third grade students are asked to name objects in pictures for the articulation test. Students are asked to talk about a picture or a scrambled outline. Students are asked to read aloud. Their responses are rated along five-point rating scales.

Attitudes were measured using McCrosky's Personal Report of Communication Fear (grades 3 and 6) and Personal Report of Communication Anxiety (grades 9 and 12).

NORM/CRITERION DATA: The instruments were field tested on 168 to 251 students per grade.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: Content was based on province developed objectives. Developers used standard procedures to identify difficulty and discrimination power of multiple choice items.

Construct and Other Empirical Studies: Factor analysis was used to substantiate structure of content domain. Results were used to revise the tests slightly.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Inter-rater reliability ranged from .75 to .83 using a 0–1 rating scale during pilot test.

Internal Consistency: Average inter-item correlations for listening tests ranged from .43 to .65. Average inter-item correlations for speaking (written) tests ranged from .39 to .64.

EVALUATIVE REACTIONS

Practicality: Written tests are very easy to administer. The oral speaking test requires small group administration and trained raters.

Validity For Specific Purposes and Populations: Evidence indicates that validity is adequate.

Reliability: Evidence indicates test reliability is fair.

Overall Adequacy: The tests cover a very broad range of skills. The tests need further development of inter-rater reliability.

MATERIALS REVIEWED: Plattor, E.; Unruh, W. R.; Muir, L.; and Loose, K. D. *Test Development for Assessing Achievement in Listening and Speaking*. Edmonton, Alberta: Alberta Education, 1978.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

21. Massachusetts Assessment of Basic Skills Listening Test

AGE RANGE: Grades 7–12.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: Not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Approximately thirty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: This test measures eleven basic listening skills that encompass understanding and using what is heard. It includes six stimuli: a new study, a commercial, a telephone conversation, a teacher's announcement, a public service announcement, and a conversation that takes place during an emergency. Each passage is brief, uses simple vocabulary and reflects common listening experiences. The test is composed of a total of twenty-two multiple-choice items. All materials including instructions and all response options are tape recorded.

NORM/CRITERION DATA: Determination of mastery level performance is not stated. A statewide survey was conducted involving 2,207 students from forty-nine schools in Massachusetts.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: Items were reviewed by a panel of judges. Item difficulties and discrimination are reported.

Construct and Other Empirical Studies: Field testing suggested that the test is not biased with respect to ethnic minorities.

RELIABILITY

Alternate Forms: Four alternate forms were developed. Difficulty levels are approximately equal. Mastery certification decisions (for assumed cut-off scores) agreed 85–90 percent.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Average inter-item correlations ranged from .64 to .86 for the four forms.

EVALUATIVE REACTIONS

Practicality: The test is easily administered, particularly since all materials are tape recorded. Test booklets are well formatted.

Validity For Specific Purposes and Populations: The test appears to be highly valid for stated objectives, with the exception that conversational listening is only obliquely reflected. The test is superior to most listening tests for assessment of life-role minimum competencies.

Reliability: The test is relatively short (twenty-two items), and this may limit reliability. A major advantage is the availability of equivalent forms for purposes of retesting those who are remediated following unsatisfactory initial testing.

Overall Adequacy: The test samples a variety of important listening situations and skills. It is not confounded with reading ability. The only significant drawback is the failure to test listening in an interactive context.

MATERIALS REVIEWED: Massachusetts Department of Education, Bureau of Research and Assessment. *Massachusetts Assessment of Basic Skills 1979–80 Development Report: Listening/Speaking*. Boston: author, 1980.

OTHER REFERENCES: Massachusetts Department of Education, *Assessment of Listening Skills State Test (Secondary Level)*. Boston: author, 1979

REVIEWER: Don Rubin

22. Massachusetts Assessment of Basic Skills Speaking Test

AGE RANGE: Grades 7–12.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: No materials; only record keeping costs.

TIME REQUIRED FOR ADMINISTRATION: Initial screening based on typical performance. One-on-one assessment less than twenty minutes per student.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: For initial screening, two classroom teachers rate a student's typical communication behavior along four dimensons: (1) delivery, (2) organization, (3) content, and (4) language. Each criterion is rated on a four-interval scale ranging from inadequate to superior. Those who do not pass the initial screening engage in four communication tasks with a single administrator/rater. The tasks are describing an activity, simulating an emergency telephone call, explaining a cooking procedure, and simulating a persuasive conversation with a school principal. The rater scores students' performances on the four tasks in the manner used for initial screening.

NORM/CRITERION DATA: The manner of determining mastery and performance level was not explained. A statewide survey was conducted involving 691 students in forty-nine schools in Massachusetts.

VALIDITY

Predictive: No information provided.

Concurrent: Scores derived from initial teacher screenings were compared with second phase communication task performance scores. Means for the two types of assessment were equal. Ratings on individual criteria were almost always within one point of each other.

Content and Item Selection: Criteria and tasks were reviewed by a panel of experts.

Construct and Other Empirical Studies: Survey data indicate some possibility of racial/ethnic bias.

RELIABILITY

Alternate Forms: Four forms of the one-on-one communication tasks were developed. All forms displayed nearly equal means. For probable cut-off scores, mastery certification decisions agreed over 90 percent, except for a narrow middle range of scores.

Test-Retest: No information provided.

Scoring: Approximately 95 percent of teacher ratings of typical communication were either equal or adjacent. Language arts and content area teachers did not differ. A subsample of one-on-one ratings were rescored with 75 percent identical rating. Some evidence of differential leniency emerged.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: Initial teacher screening considerably reduces the burden of more focused assessment. The use of a single rater/administrator in second phase testing also limits personnel demands.

Validity For Specific Purposes and Populations: The test samples varied communication contexts. It includes naturalistic observation. It uses broad evaluation criteria that encompass a broad range of skills. The question of racial/ethnic bias is undergoing additional inquiry.

Reliability: Teacher expectations are likely to play a role in initial screening, perhaps contaminating ratings with general ability. The use of only one rater in second phase also seems problematic. This issue is undergoing additional inquiry. Equivalent forms of the communication tasks are a major advantage.

Overall Adequacy: There is no guarantee that teacher ratings do, in fact, reflect only communication skills. Rubrics for criteria identify functional skills, but criteria names appear formal and absolute. The one-on-one communication tasks are fairly artificial.

MATERIALS REVIEWED: Massachusetts Department of Education, Bureau of Research and Assessment. *Massachusetts Assessment of Basic Skills 1979–1980 Development Report: Listening, Speaking*. Boston: author, 1980.

OTHER REFERENCES: Massachusetts Department of Education. *Assessment of Speaking Skills State Test (Secondary Level)*. Boston: author, 1979.

REVIEWER: Don Rubin

23. Measure of Communication Competence

AGE RANGE: Ages 2½ to 4.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: Not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Approximately fifteen minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Children are given a series of fourteen probes which form an informal interview. Responses are judged as appropriate or inappropriate. The probes and scoring guides are provided. There are two probes for each of seven modes or functions of communication. The modes are (1) contactive, (2) conversative, (3) descriptive, (4) directive, (5) explanative, (6) narrative, and (7) persuasive. These modes represent a developmental continuum of communication competencies.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: Modes were validated by F. Williams and R. Naremore, On the Functional Analysis of Social Class Differences in Codes of Speech. *Speech Monographs*, 1969 *36*, 77–102.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Inter-rater agreement was 78 percent and 81 percent.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: This approach is feasible but requires one-on-one assessment and trained scorers.

Validity For Specific Purposes and Populations: Evidence indicates that test validity is good.

Reliability: Evidence indicates that test reliability is good.

Overall Adequacy: This type of assessment measures basic, functional communication competencies. It could be adapted for older children. However, it would be harder to identify a continuum of higher level competencies.

MATERIALS REVIEWED: Riccillo, S. C. "Children's Speech and Communicative Competence." Unpublished doctoral dissertation, University of Denver, 1974.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

24. Metropolitan Achievement Tests; Listening Comprehension

AGE RANGE: Kindergarten–grade 4.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: $16.25 per thirty-five copies.

TIME REQUIRED FOR ADMINISTRATION: Approximately twenty to twenty-five minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: There is a listening comprehension component in the language subtest in four levels of the test battery: Primer, Primary 1, Primary 2, and Elementary. Each subtest has two forms. There are twenty-one to thirty listening items in the language subtest. The student is asked to pick the picture, out of four choices, that answers a question about a sentence or several sentences that the administrator then reads aloud.

NORM/CRITERION DATA: The test was normed with a stratified national sample of students. Over 550,000 students participated. Data was collected in the fall and spring.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: Test items are based on textbooks and curriculum objectives that are commonly used. Items were reviewed for bias by experts. Standard item analysis were conducted.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: No information provided.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Average inter-item correlations are available for the language subtest as a whole, which includes the listening comprehension component. For example, the reliability of the language subtest at grade 6 was .92.

EVALUATIVE REACTIONS

Practicality: The test is group administered. The test requires that the administrator dictate the items but the directions are very clear.

Validity For Specific Purposes and Populations: Although little validity related evidence is provided, the test covers standard areas of the school curriculum.

Reliability: The overall battery and major subtests are highly reliable. No evidence is given regarding the listening comprehension component.

Overall Adequacy: The test covers the basic elements of listening comprehension. However, it does not represent a breadth of listening material.

MATERIALS REVIEWED: Prescott, G. A.; Balow, I. H.; Hogan, T. P.; and Farr, R. C. *Metropolitan Achievement Tests*. New York: The Psychological Corporation, 1978.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

25. Michigan Educational Assessment Program; Listening Test

AGE RANGE: Grades 4, 7, and 10.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Approximately forty-five minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test measures several objectives of critical listening, including main idea, summary, purpose, recall of details, cause and effect, inference about the speaker or people described by the speaker, fact or opinion, story line or sequence. Several stories are read aloud from a tape recording. After each story several multiple-choice questions and response options are read aloud. The student has the questions and response options in a test booklet and marks the best answer on an answer sheet.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: No manual is included. Tests are easy to administer.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: The test measures a variety of listening skills.

MATERIALS REVIEWED: Michigan Department of Education. *Listening Test, Grades 4, 7, 10*, booklets and scripts. Lansing: author, 1978–1979.

OTHER REFERENCES: Michigan Department of Education. *Minimal Performance Objectives*. Lansing: author, 1978–1979.

REVIEWER: Nancy Mead

26. National Assessment of Educational Progress Pilot Test of Speaking and Listening

AGE RANGE: Age 17.

SKILLS TESTED:
- X Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- X Subskill or Attitude: communication

COST: Not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Fifty minutes per booklet, five booklets. None of the booklets encompasses all of the objectives.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test includes indirect measures of speaking skills in the ares of informing, controlling, and expressing feelings. The test uses a multiple-choice format to query appropriate responses to brief scenarios. Multiple-choice questions about ritualizing cover both speaking and listening skills. Tape recorded stimuli, instructions, and multiple-choice questions are used to measure listening/recognizing in the informing, controlling, and expressing feelings functions. Attitudes toward communication are measured by Likert scales.

NORM/CRITERION DATA: Test characteristics were analyzed using 693 students representing a variety of geographic regions, types of communities, and ethnic and socioeconomic backgrounds.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: Item generating matrix and objectives were based on a theory of functional communication competence. Objectives and items were developed by panels of experts and were reviewed by a minority review panel. Method of administration avoided contamination with reading ability and prior knowledge of subject matter.

Construct and Other Empirical Studies: No significant correlations between listening and speaking subtests were reported. Inconsistent correlations between subtests within speaking and listening were reported. The relationships within speaking were stronger than within listening. Inconsistent correlations between communication knowledge and attitudes were reported.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Scoring is objective, however, item analyses revealed a lack of consistent response patterns for some items and some distractors.

Internal Consistency: Average inter-item correlations were .78 for informing/speaking items, .72 for ritualizing items, .78 for communication attitudes, .66 for controlling/speaking, .63 for informing/listening, .59 for controlling/listening. Correlations were very low for speaking and listening questions in the area of expressing feelings. There were no measures of overall test reliability.

EVALUATIVE REACTIONS

Practicality: The entire test battery would require lengthy administration. Tape recorded instructions and multiple-choice response format eases administration and scoring.

Validity For Specific Purposes and Populations: There is no evidence that the indirect test of speaking knowledge predicts speaking performance. Factor analysis of communication attitudes is difficult to interpret. Some results suggest depressed, though not necessarily biased, performance for minority students. The content validity is high.

Reliability: The reliability varied considerably among content areas. Results of item analysis may be used to improve both reliability and validity.

Overall Adequacy: The test stems from a strong conceptual framework. Indirect testing of communication skills may be an untenable technique, however, since it is difficult to adequately define communication contexts and associated contingencies. Listening subtests show promise since they utilize oral language and are constructed to be uncontaminated by extraneous factors like reading ability and prior subject matter knowledge.

MATERIALS REVIEWED: Mead, N. A. "The Development of an Instrument for Assessing Functional Communication Competence of Seventeen-Year-Olds." Unpublished dissertation, University of Denver, 1977.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

27. New York State Regents Comprehensive Examination in English; Listening Section

AGE RANGE: Grade 12.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: Not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Approximately fifteen minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: A single passage is read aloud twice. The test includes ten multiple-choice items emphasizing recall, purpose, and some inference.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is easily administered and scored.

Validity For Specific Purposes and Populations: The test samples a restricted range of communication contexts. The responses are confounded with reading ability.

Reliability: A single, short passage with a ten-item test is not likely to yield reliable scores. Moreover, inconsistencies in administration due to reading of passage contribute to measurement error.

Overall Adequacy: There is little basis for interpreting scores as indices of listening achievement. The test may, however, have some value as a component of a comprehensive English test.

MATERIALS REVIEWED: University of the State of New York. *Regents High School Examination: Comprehensive Examination in English*. Albany, NY: author, 1980.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

28. New York Statewide Achievement Examination in English

AGE RANGE: Grade 12.

SKILLS TESTED:
- X Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Speaking: four minutes per student; listening: fifteen minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: For the speaking test, students present a three minute monologue on one of three supplied topics. Two minutes are allowed for preparation. Assessment takes place in class. A single rater assigns a general impression mark of 0 (nonperformance) to 10 (strong). This mark is based on (1) content, including statement of topic, focus, and support, (2) organization, including coherence and clarity, and (3) voice and articulation including volume but excluding regionalism and accent. For the listening test, a passage is read aloud. The test includes ten multiple-choice items emphasizing recall and inference.

NORM/CRITERION DATA: The test is intended for non-academically tracked students.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No informaion provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is relatively easy to administer and score. It requires little time.

Validity For Specific Purposes and Populations: The test measures a limited range of speaking skills. The speaking task is highly artificial, lacking context for communication. The criteria seek to avoid bias against nonstandard dialect speakers, but the task may be inherently biased.

Reliability: A single rater, presumably a classroom teacher, is likely to introduce rating error. However, a scoring guide with sample grade transcripts is supplied. Some provision is made for consistency of administration. Topics may affect scores.

Overall Adequacy: The test is not functionally oriented because it lacks context for communication. Speaking task and criteria taps a limited range of skills.

MATERIALS REVIEWED: University of the State of New York. *Statewide Achievement Examination in English*. Albany, NY: author, 1980.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

29. Oliphant Tests; Auditory Synthesizing Test and Auditory Discrimination Memory Test

AGE RANGE: Ages 7–14.

SKILLS TESTED:
- ___ Speaking
- ___ Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- X Subskill or Attitude: auditory memory

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: In the auditory synthesizing test, children are presented with individual sounds which they must hold in memory to form words composed of those sounds. In the discrimination test, children hear two words that sound alike. One of the two words is spoken again by the test administrator. Children must select which word was spoken twice.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is useful for children suffering from auditory problems.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: The test is quite narrow in its focus. The manual does not provide sufficient instruction for the test administrator.

MATERIALS REVIEWED: *Oliphant Tests: Auditory Synthesizing Test and Auditory Discrimination Memory Test*. Cambridge, MA: Educators Publishing Service.

OTHER REFERENCES: None.

REVIEWER: John Daly

30. Oral Language Evaluation

AGE RANGE: Not specified, but appears appropriate for elementary level.

SKILLS TESTED:
- X Speaking
- X Listening
- X Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- X Subskill or Attitude: assesses child's primary language, identifies children who need training in English

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Part 1 not specified; part 2, two minutes per child; part 3, ten minutes per child.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is designed to identify, assess, diagnose, and prescribe the oral language ability of English and Spanish speaking children. Part 1 identifies children who may need training in a second language. This is assessed by teacher observation and information from school records. The child is judged to need English as a second language (ESL) training if the primary language in the home is not English, if the student most often speaks a language other than English, or if this child's first acquired language is not English. Part 2 assesses the child's primary language and determines if additional testing in oral language ability is necessary. The test administrator presents four pictures and encourages the child to discuss them. The test administrator uses a "Six Level Language Continuum" to determine if further testing is necessary. Examples of responses at various levels are provided. The same procedure is followed for Part 3 except the child's discussion of the pictures is tape recorded. Transcripts of these tapes are analyzed using the "Six Level Language Continuum." Part 4 provides instructional activities for children at each level on the continuum.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is administered by teacher; no training is necessary. Materials are provided in the teacher's manual.

Validity For Specific Purposes and Populations: Although no validity information is provided, the test may give a reasonable estimate of the child's "primary" language skills.

Reliability: Inadequate information to judge.

Overall Adequacy: These activities would be helpful to teachers working with English as a second language children, but they do not constitute a test of communication skill. There is a serious lack of rigorous evaluation.

MATERIALS REVIEWED: Silvaroli, N. J.; Skinner, J. T.; and Maynes, Jr., J. O. *Oral Language Evaluation*. St. Paul, MN: EMC Corporation, 1977.

OTHER REFERENCES: None.

REVIEWER: Janice Patterson

31. Profile of Nonverbal Sensitivity

AGE RANGE: Secondary.

SKILLS TESTED:
- ___ Speaking
- ___ Listening
- ___ Interaction
- ___ Visual Encoding
- X Visual Decoding
- ___ Subskill or Attitude

COST: Videotape and 16mm film not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Approximately forty-five minutes for full 220 item test; short versions are available.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The full test includes 220 items composed of eleven channels. The channels include facial expression, body, face and body, electronically altered voice (free of language content), and various combinations thereof. Stimuli are presented on film or videotape with pauses for identifying items and responding. Answer sheet is a two option multiple-choice format. Response options are behavioral descriptions such as "asking forgiveness" or "talking about one's divorce."

NORM/CRITERION DATA: The main standardization group was composed of 492 high school students of average intelligence from three geographical regions. Other norming studies were based on various adult, international and impaired populations.

VALIDITY

Predictive: No information provided.

Concurrent: For high school students, test scores were slightly correlated with intelligence and SAT scores. Moderate correlations were found between test scores and various other tests of nonverbal decoding including *Communication of Affect Receiving Ability Test* and *Social Interpretations Test*. Zero, low, or moderate correlations emerged between test scores and several measures of psychological traits such as dogmatism and teacher aptitude. Test scores correlated strongly with other ratings of social competence and sensitivity.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: Test stimuli accounted for a wide range of nonverbal information including movement, full body posture, and voice. A large number of empirical studies demonstrated that the instrument can discriminate among various occupational and social categories of subjects. Training improved measured nonverbal sensitivity. Other studies found no ethnic bias, but some cross-cultural bias, favoring Americans as opposed to other nationalities.

RELIABILITY

Alternate Forms: Various short forms of the instrument have been developed. Accuracy was significantly greater on the full film version than on a still photograph version. Low or moderate correlations were found for other shortened versions and the full test.

Test-Retest: For 293 subjects, both adult and high school, with second exposure from ten days to eight weeks after first, pooled reliability was .69.

Scoring: Not applicable.

Internal Consistency: Average inter-item correlation was .86.

EVALUATIVE REACTIONS

Practicality: The test is well packaged for ease of administration and scoring. It requires a full class period.

Validity For Specific Purposes and Populations: Stimuli cover a wide range of nonverbal signals. The major threats to validity pertain to the test response mode. Students must be able to read behavioral descriptions. Moreover, the meaning of some of the descriptors may be variable or unfamiliar for many students. With only two options for each item, item difficulty may be too low for adequate discrimination in many cases.

Reliability: Use of short versions is questionable, but reliability is otherwise quite respectable for a communication decoding task involving the recognition of affect.

Overall Adequacy: Analysis of children's understanding of the response terms is necessary to interpret test results. The test stimulus appears to have high ecological validity for the range of nonverbal sensitivity measured.

MATERIALS REVIEWED: Rosenthal, R.; Hall, J. A.; DiMatteo, M. R.; Rogers, P. L.; and Archer, D. *Sensitivity to Nonverbal Communication: The PONS Test*. Baltimore: Johns Hopkins, 1979.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

32. PRI Reading Systems, Oral Language Skill Clusters

AGE RANGE: Kindergarten–grade 3.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- X Visual Decoding
- ___ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Fifty to seventy minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is composed of subscales that assess (1) sound segmentation, (2) vocabulary, (3) syntax, (4) literal meaning, and (5) inferred meaning. The test is basically a phonetic discrimination and listening test. There are two levels of the test. Level A covers kindergarten and grade 1. Level B covers grades 2 and 3. Multiple-choice questions are orally presented by the teacher.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test seems suited to the classroom teacher. The success of the test depends on appropriate administration of the test by the teacher.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: The test is very limited in what it assesses. Results could be confounded with intelligence, teacher delivery, and other constructs.

MATERIALS REVIEWED: *PBI Reading Systems*. New York: McGraw Hill, 1980.

OTHER REFERENCES: None.

REVIEWER: John Daly

33. SRA Achievement Series

AGE RANGE: Kindergarten–grade 3.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- X Subskill or Attitude: auditory discrimination

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Auditory recognition, twenty or twenty-five minutes; listening comprehension, twenty-five minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: There are three levels of the test: A (kindergarten–grade 1), B (grades 1–2), and C (grades 2–3). Each level has two forms. For the auditory discrimination test (Levels A–B), the test administrator reads two words and the student must answer whether the words are alike or different in one of the following ways: (1) beginning sounds, (2) vowels, (3) ending sounds, and (4) beginning and ending sounds.

In the listening comprehension test (Levels A–C) the student is required to identify the correct illustration for a word or situation read aloud by the test administrator. Skills include: (1) identifying a picture specified by oral directions, (2) identifying a picture of a detail in a sentence or story given orally, (3) identifying a picture of the main idea of a sentence or story given orally, (4) identifying a picture of a relationship among events in a story given orally (such as sequence or cause), and (5) identifying a picture of a conclusion based on material given orally.

NORM/CRITERION DATA: The tests were normed with at least 3,000 students per grade per form in the first standardizing. The tests were standardized a second time with a smaller group.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: Extensive content validity and item selection procedures were implemented. Items were reviewed for bias. Statistical tests of bias were conducted.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: No information provided.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Based on data from Form 1, average inter-item correlations for the auditory discrimination test ranged from .79 to .89. Average inter-item correlations for the listening comprehension test ranged from .55 to .80.

EVALUATIVE REACTIONS

Practicality: The tests provide complete, easy to follow directions.

Validity For Specific Purposes and Populations: Evidence indicates the test validity is good.

Reliability: Evidence indicates the reliability of the auditory discrimination test is good; the reliability of the listening comprehension test is fair.

Overall Adequacy: These tests are designed to measure reading readiness. The tests are not as good for measures of communication.

MATERIALS REVIEWED: *SRA Achievement Series*. Chicago, IL: Science Research Associates, 1979.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

34. Sequential Tests of Educational Progress: Listening

AGE RANGE: Grades 3–12.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: $12.

TIME REQUIRED FOR ADMINISTRATION: Twenty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test has six levels; one for each of six age groups. Each level has two forms (X and Y). The teacher reads a short passage to students, students then answer multiple-choice questions about the passage. Direction following is also included in the test. Teachers read instructions to students who work on the dictated problem using a worksheet.

NORM/CRITERION DATA: Specific studies were not summarized. Norm, classification, and percentile values are in the process of being derived.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: No information provided.

Test-Retest: No information provided

Scoring: Not applicable.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test appears to have very good potential.

Validity For Specific Purposes and Populations: The test meets the specific goals stated in the justification.

Reliability: Inadequate information to judge.

Overall Adequacy: Inadequate information to judge. Studies are likely in progress. This test may become a standard listening test.

MATERIALS REVIEWED: Educational Testing Service. *Sequential Test of Educational Progress*. Reading, MA: Addison-Wesley, 1979.

OTHER REFERENCES: None.

REVIEWER: John Daly

35. Stanford Achievement Tests; Listening Comprehension

AGE RANGE: Grades 1–6.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: $20.50 per thirty-five booklets.

TIME REQUIRED FOR ADMINISTRATION: Approximately twenty-five to thirty-five minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: A listening comprehension component is included in five levels of the test battery: Primary I, Primary II, Primary III, Intermediate I, and Intermediate II. Each test has two forms (A and B). There are twenty-six items in the lowest level and fifty items in all other levels. The student hears a wide variety of passages dictated by the administrator and picks the picture or answer, out of four choices, that best answers a question about the passage. The questions measure central focus, specific meanings, implied meanings, perceptions of concepts and relations, and identification of inferences.

NORM/CRITERION DATA: The test was normed on a national sample of students that represented various locations and types of communities. Over 225,000 students were involved. The study included fall and spring administrations.

VALIDITY

Predictive: No information provided.

Concurren.. Correlations between the listening test and the other tests in the battery and with the Otis-Lennon Mental ability were generally high, ranging in the .50s to the .80s, indicating that all tests are operating in the same general domain. Correlations were about as high with math as with reading.

Content and Item Selection: The test was based on commonly used textbooks and curriculum objectives. Items were reviewed for bias by experts. Standard item analyses were performed.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: No information provided.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Split-half reliabilities ranged from .86 and .88 and average inter-item correlations ranged from .85 to .89 for three grades about which the reviewer had information.

EVALUATIVE REACTIONS

Practicality: The test is group administered. The listening comprehension items must be dictated by the administrator but the instructions are very clear.

Validity For Specific Purposes and Populations: Although not much evidence of validity is provided, the test covers the areas commonly included in curricula.

Reliability: The evidence indicates the reliability of the listening component is very good.

Overall Adequacy: The test attempts to cover a variety of listening passages and types of questions. Some questions, however, appear to test thinking skills or vocabulary knowledge rather than comprehension.

MATERIALS REVIEWED: Madden, R.; Gardner, E. F.; Rudman, H. P.; Karlsen, B.; and Merwin, J. C. *Stanford Achievement Test*. New York: Harcourt Brace Jovanovich, Inc., 1973.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

36. Stanford Early School Achievement Test

AGE RANGE: Kindergarten–grade 1.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Approximately ninety minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is designed to assess children's cognitive abilities. The test consists of four parts: (1) the environment, (2) mathematics, (3) letters, and (4) sounds and aural comprehension. This is a group administered test where children write in answer booklets to indicate their responses. Subscores are available in each of the four parts.

NORM/CRITERION DATA: Norms were determined from responses of children from twenty-five states. The final norm sample consisted of 8,310 kindergarten and 11,106 first graders. Census, "size of city" data, and intelligence scores were used in selecting these children.

VALIDITY

Predictive: No information provided.

Concurrent: The test correlated .74 with the *Otis-Lennon Mental Ability Test* for 11,106 first graders.

Content and Item Selection: Original questions were given to 3,100 first grade children in ten school districts. The best items were selected from those forms.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: The split-half reliability coefficients ranged from .76 to .85 for kindergarten and .77 to .89 for first grade.

EVALUATIVE REACTIONS

Practicality: Children are tested in groups of seven to fifteen. The test may be impractical for large scale testing. No special training is necessary; the test may be administered by teacher.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: This is not a test of speech communication, rather a test of cognitive abilities. Instruction manual includes classroom activities.

MATERIALS REVIEWED: Madden, R., and Gardner, E. F. *Stanford Early School Achievement Test*. New York: Harcourt Brace Jovanovich, Inc., 1969.

OTHER REFERENCES: None.

REVIEWER: Janice Patterson

37. Situational Language Tasks

AGE RANGE: Grades 1–3.

SKILLS TESTED:
- X Speaking
- X Listening
- X Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: User produces materials and scores.

TIME REQUIRED FOR ADMINISTRATION: Fifteen minutes per session; three sessions.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is divided into three parts. The first section is designed to assess classroom teacher-child and child-child interaction. The teacher and the whole class discuss an assortment of common objects. The teacher is instructed to use these materials to elicit conversation from the children. The session is tape recorded and lasts fifteen minutes. In the second section, conducted on a different day, three children and a teacher discuss cartoon-like pictures. The teacher shows one picture at a time and asks the children a series of structured questions such as, "What is happening in this picture?" In the second phase of section 2, the teacher shows two cards and asks, "What picture comes first?" In the final five minute phase of section 2, the teacher presents three cards and asks, "What story does this picture tell?" All conversations are tape recorded. The third section of the test occurs immediately following section 2. The teacher tells the small group of children that she has work to do but that they may stay and discuss the cards. The tape recorder continues to run and the children's resulting conversations are recorded for analysis. Transcripts of the children's speech are analyzed in four major areas: (1) type-token ratio (ratio of the number of different words to the total number of words), (2) verb tense diversity, (3) vocabulary diversity, and (4) average number of words per child.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: The tasks were selected to reflect children's natural language in the classroom.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: The consensus method of reliability was used for the tape transcriptions. Five researchers coded the transcripts; the reliability ranged from 89 percent to 100 percent.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is not practical for large scale assessment because children work in small groups and scoring and analysis are complicated. Resulting data would not be available to teachers in a useful form.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: The test is a poor assessment tool for measuring communication ability. Although it focuses on children's speech, the serious issues of validity and reliability are unaddressed. It also seems to tap other skills such as sequencing in session 2. The scoring system is complicated and cumbersome for teachers.

MATERIALS REVIEWED: The University of Arizona, College of Education. *Use of Situational Language Tasks in an Intra-TEEM and TEEM Versus Comparison Evaluation*. Tucson, AZ: author, 1976.

OTHER REFERENCES: None.

REVIEWER: Janice Patterson

38. Speech in the Classroom: Assessment Instruments of Speaking Skills

AGE RANGE: Grades 1–12.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- X Subskill or Attitude: speaking experience and attitudes

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Assessment of Speaking, approximately five minutes; Inventory of Experiences, approximately five minutes; and Summary of Attitudes, approximately five minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test consists of three parts. The first is an assessment of speaking skills. This is a one-on-one assessment of the student with the administrator. The student chooses a picture to tell a story about and then performance is rated on a 1–4 point scale. The second part is an inventory of classroom speaking experiences. This is a paper and pencil test. The test has two levels, one for grades 1–6 and one for grades 4–12. The test asks students fifteen or twenty-five questions about speaking experience in the classroom. The test asks the teacher fifteen to twenty-five related questions about speaking activities in the classroom. The third part is a survey of attitudes toward classroom speech situations. This is a paper and pencil test. The test has two levels, one for grades 1–6 and one for grades 4–12. The test asks the student twelve or twenty questions about attitudes toward self speaking in class.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The assessment of speaking skills is feasible but the test requires one-on-one assessment and trained scorers. The other tests are easy to administer.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Inadequate information to judge.

Overall Adequacy: The test is in the developmental stages. It needs more testing and documentation.

MATERIALS REVIEWED: Kozpol, S., and Cercone, K. *Speech in the Classroom: Assessment Instruments*. Harrisburg, PA: Pennsylvania Department of Education, 1980.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

39. Test of Adolescent Language

AGE RANGE: Ages 11–18.

SKILLS TESTED:
- X Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Approximately $75.

TIME REQUIRED FOR ADMINISTRATION: One to three hours (test is open ended).

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is designed to identify language proficiency, to identify strengths and weaknesses in various dimensions of language, and to measure progress. It includes eight substates covering reading, writing, speaking, and listening. Four are related to speaking and listening. For Listening/Vocabulary, a word is read aloud and students identify two pictures which relate to the word. For Listening/Grammar, three sentences are read aloud and students identify two that express the same thought. For Speaking/Vocabulary, a word is read aloud and the student uses the word correctly in a meaningful sentence. For Speaking/Grammar, a sentence is read aloud and students repeat it aloud.

NORM/CRITERION DATA: The test was normed on 2,723 students in seventeen states between ages 11 and 18.

VALIDITY

Predictive: No information provided.

Concurrent: With thirty-two subjects, total composite listening correlated .51 with *Peabody Picture Vocabulary Test*.

Total composite speaking correlated .60 with memory for related syllables from *Detroit Tests of Learning Aptitude*.

Content and Item Selection: Test developers used standard procedures to identify item difficulty and discrimination power.

Construct and Other Empirical Studies: Various studies supported hypotheses of age differentiation, subtest interrelationship, group differentiation, and relationship with tests of intelligence.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: Correlations ranged from .74 to .85, with higher correlations for composites.

Scoring: Inter-rater reliability for speaking/vocabulary was .96.

Internal Consistency: Average inter-item correlations for subtests ranged from .60 to .90 with higher concentrations for composites.

EVALUATIVE REACTIONS

Practicality: The speaking tests require individual administration and trained scorers but are straightforward. Listening tests are easy to administer.

Validity For Specific Purposes and Populations: Evidence indicates test validity is very good.

Reliability: Evidence indicates test reliability is very good.

Overall Adequacy: The measures focus on very narrow subskills.

MATERIALS REVIEWED: Hammill, D. D.; Brown, V. L.; Larsen, S. C.; and Widerhoit, J. L. *Test of Adolescent Language*. Austin, TX: PRO-ED, 1980.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

40. Test of Listening Accuracy in Children

AGE RANGE: Kindergarten–grade 6.

SKILLS TESTED:
- __ Speaking
- X Listening
- __ Interaction
- __ Visual Encoding
- X Visual Decoding
- __ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Two tests are available. One is for individual testing and the other for group testing. Children are presented with picture pairs and then hear one of the pictures named. They are to identify the appropriate picture.

NORM/CRITERION DATA: Some norms are reported and qualitative classes (average/superior, etc.) are provided, but no research base for the categorical assignments is provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Internal consistency correlations ranged from .75 to .95 depending on groups and scales.

EVALUATIVE REACTIONS

Practicality: The ability to use this test with either groups or individuals makes it attractive. It seems usable by classroom teachers.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Evidence indicate test reliability is good.

Overall Adequacy: The test may be useful for teachers but suffers from insufficient normative, empirical, and user-oriented data.

MATERIALS REVIEWED: Mecham, M. J.; Jex, J. L.; and Jones, J. D. *Test of Listening Accuracy in Children, Manuals for Individual and Group Testing*. Salt Lake City: Communication Associates, Inc.

OTHER REFERENCES: None.

REVIEWER: John Daly

41. Torrance Tests of Creative Thinking; Verbal Test

AGE RANGE: Kindergarten–grade 3.

SKILLS TESTED:
- X Speaking
- ___ Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: $8.50 for twenty-five booklets; $1.45 for scoring each student.

TIME REQUIRED FOR ADMINISTRATION: Approximately forty-five minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The verbal test measures creative thinking using words. The test has two forms (A and B). For students from kindergarten through grade 3, the test is administered individually and students give responses orally. At older ages the test is group administered and individuals respond in writing. The test includes seven tasks which require creative responses, for example, listing possible causes for events shown in a picture. Responses are evaluated in terms of fluency, originality and in some cases flexibility. The fluency score is primarily the number of relevant responses. An optional scoring for elaboration is mentioned but no scoring guides are provided.

NORM/CRITERION DATA: Some norm data are provided but the technical manual was not available for review.

VALIDITY

Predictive: Some predictive data are provided but the technical manual was not available for review.

Concurrent: No information available.

Content and Item Selection: No information available.

Construct and Other Empirical Studies: No information available.

RELIABILITY

Alternate Forms: No information available.

Test-Retest: No information available.

Scoring: Correlation between scores of trained and untrained scores ranged from .86 to .96. More information is available in the technical manual.

Internal Consistency: No information available.

EVALUATIVE REACTIONS

Practicality: The test requires very lengthy individual testing. Maintaining the attention span of young children might be difficult. Minimal guidance is given for translating tasks into language for young children.

Validity For Specific Purposes and Populations: Inadequate information to judge.

Reliability: Preliminary evidence indicates high scorer reliability.

Overall Adequacy: The only measure of oral ability in this test is one of quantity.

MATERIALS REVIEWED: Torrance, E. P. *Torrance Tests of Creative Thinking*. Bensenville, IL: Scholastic Testing Service, 1966.

OTHER REFERENCES: None.

REVIEWER: Nancy Mead

42. Utah Test of Language Development (Direct Test Version)

AGE RANGE: Preschool–grade 9 (ages 2–14).

SKILLS TESTED:
- X Speaking
- X Listening
- X Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not specified.

TIME REQUIRED FOR ADMINISTRATION: Twenty to forty minutes.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test is designed to assess language production and comprehension skills. Test items include responding to instructions, naming objects, repeating digital span (forewards and backwards), indicating receptive vocabulary, drawing simple shapes, writing numbers and letters (both manuscript and cursive), telling a story, and reading on a primer level. Scoring is for total test performance and does not provide subtest scores. Two forms are available: the direct-test version and the informant-interview version (available through the American Guidance Association).

NORM/CRITERION DATA: Norms are based on 393 children in twenty-three states. These data were combined with a Utah sample of 273 children, judged representative of a normal population. Later, data were collected on 989 kindergarten children including minorities. The norms provide language-age equivalents.

VALIDITY

Predictive: No information provided.

Concurrent: The test correlated .72 and .81 with the *Verbal Language Development Scale*, .53 with the *Mean Length of Utterances*, and .87 and .91 with the *Illinois Test of Psycholinguistic Abilities*.

Content and Item Selection: Items were selected from standard sources.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: The two forms, direct test and the informant interview form, correlated .81 with a time interval of approximately three weeks.

Test-Retest: See above.

Scoring: No information provided.

Internal Consistency: Split-half correlation was .94.

EVALUATIVE REACTIONS

Practicality: The test is administered individually, thus, it is not practical for large scale testing. No specific training is needed.

Validity For Specific Purposes and Populations: The stated purpose of assessing broad language skills is violated by inclusion of items on reading and small motor skills (writing and drawing) and items requiring memory of a digital span (often seen on intelligence tests).

Reliability: Evidence indicates test reliability is very good.

Overall Adequacy: The test is a poor measure of communication because many items call for proficiencies other than language ability.

MATERIALS REVIEWED: Mecham, M. J., and Jones, J. D. *Utah Test of Language Development Manual of Directions*. Salt Lake City: Communication Research Associates, Inc., 1978.

OTHER REFERENCES: None.

REVIEWER: Janice Patterson

43. Vermont Basic Competency Program Speaking and Listening Assessments

AGE RANGE: Kindergarten–grade 12

SKILLS TESTED:
- X Speaking
- X Listening
- X Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: No costs beyond record keeping.

TIME REQUIRED FOR ADMINISTRATION: Highly variable.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Precise implementation of this test is determined by local districts, and apparently varies across age levels. Speaking tasks include (1) giving directions, (2) extended expository, informative, or persuasive talk, (3) answering telephones and taking messages, (4) using telephones to get information and assistance, (5) introducing self and others, (6) interviewing for a job, and (7) participating in informal discussion. Listening tasks include (1) following directions, (2) retelling a narrative sequence, and (3) summarizing factual material. Several of these tasks involve simulated tasks, while informal discussion skill is assessed naturalistically over time. Evaluation criteria are not wholly specified, but apparently include accuracy, use of standard English organization, clear articulation, and other functionally related criteria.

NORM/CRITERION DATA: The percent of 12-, 14-, and 15-year-olds passing each competency are reported. These results apparently summarize scores of several thousand students throughout Vermont.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: The development of competencies (tasks) were based on input of 1,500 Vermont educators as well as extensive search of literature.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The procedures place the burden of testing and record keeping on teachers. It is difficult to check on compliance in implementation. The test would require massive in-service training, with probable beneficial effects. Ideally the procedures would include a second rater which would increase cost.

Validity For Specific Purposes and Populations: The test includes a good sampling of communication situations. The lack of precisely defined evaluation criteria makes validity difficult to judge. If the procedures are closely tied to instruction it may not be a valid measure of individual ability.

Reliability: Use of single classroom teacher rating performance without clearly defined guidelines is a major problem. Expectations and bias would likely be major factors. Also, procedures make no provision for consistency of administration.

Overall Adequacy: The program seems well motivated by a concern for functional communication competence. Use of contextually diverse tasks is especially admirable. But lack of well defined procedures compromise the value of results.

MATERIALS REVIEWED: Vermont Department of Education. *Basic Competencies: A Manual of Information and Guidelines for Teachers and Administrators, Vermont's Basic Competency Program, 1978–1979 Report, 1979–1980 Report*. Montpelier: author, 1978–79, 1979–80.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

44. Wallner Test of Listening Comprehension

AGE RANGE: Kindergarten–grade 1.

SKILLS TESTED:
- ___ Speaking
- X Listening
- ___ Interaction
- ___ Visual Encoding
- ___ Visual Decoding
- ___ Subskill or Attitude

COST: Not commercially available.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: The test has two forms (A and B). Each form consists of six passages of graded readability. Each passage is followed by seven literal comprehension and seven inferential questions. The test administrator reads the passages and questions aloud. Students pick the picture, from three choices, that best answer the question.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: One year after taking the test, 107 students were given the Stanford Achievement Test, Primary I Battery, Form W. Correlations between Forms A and B and the Stanford test was .68 and .64 respectively.

Concurrent: Correlations between forms A and B and the Listening Subtest of the Metropolitan Readiness Tests (Form B) were .59 and .60 respectively, based on a sample of 150 students. Correlations with the Metropolitan Readiness Test total score were .74 and .72.

Content and Item Selection: Passages were composed on the basis of the Dale-Chall and Spache Readability formulas. A panel of experts in reading verified the skill placement and content validity of the items.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: The two forms correlated .89 based on a sample of 140 students.

Test-Retest: No information provided.

Scoring: Not applicable.

Internal Consistency: Average inter-item correlations were .95 for Form A and .95 for Form B, based on a sample of 140 students.

EVALUATIVE REACTIONS

Practicality: No information is provided about the length of test administration. However, administration and scoring procedures are not complicated.

Validity For Specific Purposes and Populations: The test seems more closely related to general verbal ability than to listening ability *per se*. Effects of children's prior knowledge about subject matter is unknown.

Reliability: Variation in administration may adversely affect reliability. The test appears highly reliable in other respects.

Overall Adequacy: The test adopts a non-interactive definition of listening comprehension based on extended written prose rather than oral language stimuli.

MATERIALS REVIEWED: Wallner, N. K. "The Development of a Listening Comprehension Test for Kindergarten and Beginning First Grade." *Educational and Psychological Measurement* 34(1974):391–396.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

45. Westside High School Minimum Competency Test

AGE RANGE: Grade 10.

SKILLS TESTED:
- X Speaking
- __ Listening
- __ Interaction
- __ Visual Encoding
- __ Visual Decoding
- __ Subskill or Attitude

COST: Only record keeping costs.

TIME REQUIRED FOR ADMINISTRATION: Not specified.

DESCRIPTION OF TEST, PROCEDURES, ITEMS, SCORING: Students choose a topic and individually present an extended discourse to a group. A planning sheet emphasizing purpose, development, and organization is provided. The student is encouraged to rehearse the talk prior to testing. Dichotomously scored criteria include (1) introduction, (2) supporting material, (3) conclusion, (4) language including grammar and word choice improprieties, (5) volume, (6) eye contact, and (7) response to questions. Students must demonstrate mastery on all criteria for a passing score.

NORM/CRITERION DATA: No information provided.

VALIDITY

Predictive: No information provided.

Concurrent: No information provided.

Content and Item Selection: No information provided.

Construct and Other Empirical Studies: No information provided.

RELIABILITY

Alternate Forms: Not applicable.

Test-Retest: No information provided.

Scoring: No information provided.

Internal Consistency: No information provided.

EVALUATIVE REACTIONS

Practicality: The test is easily administered and scored.

Validity For Specific Purposes and Populations: The test measures a very narrow range of situations. The criteria are more formal than functional. Emphasis on language "errors" may bias the test against speakers of nonstandard dialects.

Reliability: Use of a single observer calls reliability into question. Allowing students free choice of topic may also introduce measurement error.

Overall Adequacy: The test fails to sample a spectrum of communication competencies. No provision is made for simulating a communicative context. Criteria emphasizes mechanical aspects of public speaking.

MATERIALS REVIEWED: Westside Community Schools. *Minimum Competency Packet.* Omaha, NE: author, 1979.

OTHER REFERENCES: None.

REVIEWER: Don Rubin

APPENDICES

Appendix A

Standards for Effective Oral Communication Programs

Prepared by: American Speech-Language-Hearing Association and Speech Communication Association

Adequate oral communication frequently determines an individual's educational, social, and vocational success. Yet, American education has typically neglected formal instruction in the basic skills of speaking and listening. It is important that state and local education agencies implement the most effective oral communication programs possible.

The following standards for oral communication were developed by representatives of the Speech Communication Association and the American Speech-Language-Hearing Association.

If effective oral communication programs are going to be developed, all components of the recommended standards must be considered. Implementation of these standards will facilitate development of adequate and appropriate oral communication necessary for educational, social, and vocational success.

Definition

Oral Communication: the process of interacting through heard and spoken messages in a variety of situations.

Effective oral communication is a learned behavior, involving the following processes:

1. Speaking in a variety of educational and social situations: Speaking involves, but is not limited to, arranging and producing messages through the use of voice, articulation, vocabulary, syntax, and nonverbal cues (e.g., gesture, facial expression, vocal cues) appropriate to the speaker and listeners.
2. Listening in a variety of educational and social situations: Listening involves, but is not limited to, hearing, perceiving, discriminating, interpreting, synthesizing, evaluating, organizing, and remembering information from verbal and nonverbal messages.

Basic Assumptions

1. Oral communication behaviors of students can be improved through direct instruction.
2. Oral communication instruction emphasizes the interactive nature of speaking and listening.
3. Oral communication instruction addresses the everyday communication needs of students and includes emphasis on the classroom as a practical communication environment.
4. There is a wide range of communication competence among speakers of the same language.
5. Communication competence is not dependent upon use of a particular form of language.
6. A primary goal of oral communication instruction is to increase the students' repertoire and use of effective speaking and listening behaviors.
7. Oral communication programs provide instruction based on a coordinated developmental continuum of skills, preschool through adult.
8. Oral communication skills can be enhanced by using parents, supportive personnel, and appropriate instructional technology.

An Effective Communication Program Has the Following Characteristics:

Teaching/Learning

1. The oral communication program is based on current theory and research in speech and language development, psycholinguistics, rhetorical and communication theory, communication disorders, speech science, and related fields of study.
2. Oral communication instruction is a clearly identifiable part of the curriculum.
3. Oral communication instruction is systematically related to reading and writing instruction and to instruction in the various content areas.
4. The relevant academic, personal, and social experiences of students provide core subject matter for the oral communication program.
5. Oral communication instruction provides a wide range of speaking and listening experience, in order to develop effective communication skills appropriate to:
   a. a range of situations; e.g., informal to formal, interpersonal to mass communication.
   b. a range of purposes; e.g., informing, learning, persuading, evaluating messages, facilitating social interaction, sharing feelings, imaginative and creative expression.
   c. a range of audiences; e.g., classmates, teachers,

peers, employers, family, community.

d. a range of communication forms; e.g., conversation, group discussion, interview, drama, debate, public speaking, oral interpretation.

e. a range of speaking styles; impromptu, extemporaneous, and reading from manuscript.

6. The oral communication program provides class time for systematic instruction in oral communication skills, e.g., critical listening, selecting, arranging, and presenting messages giving and receiving constructive feedback, nonverbal communication, etc.
7. The oral communication program includes development of adequate and appropriate language, articulation, voice, fluency and listening skills necessary for success in educational, career, and social situations through regular classroom instruction, cocurricular activities, and speech-language pathology and audiology services.
8. Oral communication program instruction encourages and provides appropriate opportunities for the reticent student (e.g., one who is excessively fearful in speaking situations), to participate more effectively in oral communication.

Support

1. Oral communication instruction is provided by individuals adequately trained in oral communication and/or communication disorders, as evidenced by appropriate certification.
2. Individuals responsible for oral communication instruction receive continuing education on theories, research, and instruction relevant to communication.
3. Individuals responsible for oral communication instruction participate actively in conventions, meetings, publications, and other activities of communication professionals.
4. The oral communication program includes a system for training classroom teachers to identify and refer students who do not have adequate listening and speaking skills, or are reticent, to those qualified individuals who can best meet the needs of the student through further assessment and/or instruction.
5. Teachers in all curriculum areas receive information on appropriate methods for: (a) using oral communication to facilitate instruction, and (b) using the subject matter to improve students' oral communication skills.
6. Parent and community groups are informed about and provided with appropriate materials for effective involvement in the oral communication program.
7. The oral communication program is facilitated by availability and use of appropriate instructional materials, equipment, and facilities.

Assessment and Evaluation

1. The oral communication program is based on a schoolwide assessment of the speaking and listening needs of students.
2. Speaking and listening needs of students will be determined by qualified personnel utilizing appropriate evaluation tools for the skills to be assessed, and educational levels of students being assessed.
3. Evaluation of student progress in oral communication is based upon a variety of data including observations, self-evaluations, listeners' responses to messages, and formal tests.
4. Evaluation of students' oral communication encourages, rather than discourages, students' desires to communicate by emphasizing those behaviors which students can improve, thus enhancing their ability to do so.
5. Evaluation of the total oral communication program is based on achievement of acceptable levels of oral communication skill determined by continuous monitoring of student progress in speaking and listening, use of standardized and criterion-referenced tests, audience-based rating scales, and other appropriate instruments.

Appendix B

Criteria for Evaluating Instruments and Procedures for Assessing Speaking and Listening

The following criteria may be applied to published and unpublished instruments and procedures for assessing speaking and listening skills of children and adults. The criteria are organized around (a) content considerations, which deal primarily with the *substances* of speaking and listening instruments and procedures, and (b) technical considerations, which deal with such matters as reliability, validity, and information on administration.

1. Stimulus materials should require the individual being tested to *demonstrate* skill as a speaker or listener.
2. Assessment instruments and procedures should clearly distinguish speaking and listening performance from reading and writing ability; i.e., inferences of speaking and listening competence should not be made from tests of reading and writing, and directions and responses for speaking and/or listening tests should not be mediated through reading and writing modes.
3. Assessment instruments and procedures should be free of sexual, cultural, racial, and ethnic content and/or stereotyping.
4. Assessment should confirm the presence or absence of skills, not diagnose reasons why individuals demonstrate or fail to demonstrate those skills.
5. Assessment should emphasize the application of speaking and listening skills that relate to familiar

situations; i.e., stimulus materials should refer to situations recognizable to the individual being tested and should facilitate demonstration of skills rather than demonstration of content mastery.

6. Assessment should test skills that are important for various communication settings (e.g., interpersonal, small group, public, and mass communication settings) rather than be limited to one setting.
7. Assessment should permit a range of *acceptable* responses, where such a range is appropriate.
8. Assessment should demonstrate that outcomes are more than just chance evidence; i.e., assessment should be reliable.
9. Assessment should provide results that are consistent with other evidence that might be available.
10. Assessment should have content validity.
11. Assessment procedures should be standardized and detailed enough so that individual responses will not be affected by the administrator's skills in administering the procedures.
12. Assessment procedures should approximate the recognized stress level of oral communication; they should not increase or eliminate it.
13. Assessment procedures should be practical in terms of cost and time.
14. Assessment should involve simple equipment.
15. Assessment should be suitable for the developmental level of the individual being tested.

Developed by Philip M. Backlund, Kenneth L. Brown, Joanne Gerry, and Fred E. Jandt acting as a subgroup of the Speech Communication Association's Educational Policies Board Task Force on Assessment and Testing. Approved and endorsed by the Educational Policies Board and the Administrative Committee of the Speech Communication Association.

Bibliography

Allen R. R., and Brown, K. L., eds. *Developing Communication Competence in Children*. Skokie, IL: National Textbook, 1976.

Backlund, P. "A National Survey of Elementary and Secondary Speaking and Listening Assessment." In *Perspective on Oral Communication Assessment in the 80's*, edited by R. Stiggins. Portland, OR: Northwest Regional Educational Laboratory, 1981.

Bales, R. *Interaction Process Analysis*. Cambridge: Addison-Wesley, 1953.

Barnes, D. "Situated Speech Strategies: Aspects of the Monitoring of Oracy." *Educational Research* 32(1980):123–131.

Bassett, R. E.; Whittington, N.; and Staton-Spicer, A. "The Basics in Speaking and Listening for High School Graduates: What Should Be Assessed?" *Communication Education* 27(1978):293–303.

Bazen, D. "The Place of Conversation Tests in Oral Examinations." *English in Education* 12(1978):39–50.

Becker, S. L. "The Rating of Speeches: Scale Independence." *Speech Monographs* 29(1962):38–44.

Braddock, R.; Lloyd-Jones, R.; and Shoer, L. *Research in Written Composition*. Champaign, IL: National Council of Teachers of English, 1963.

Brown, K. L.; Backlund, P.; Gurry, J.; and Jandt, F. *Assessment of Basic Speaking and Listening Skills* (2 vols.). Boston: Massachusetts Department of Education, 1979.

Buros, O. K. *The Eighth Mental Measurements Yearbook* (2 vols.). Highland Park, NJ: Gryphon, 1978.

Carroll, B. J. *Testing Communicative Performance: An Interim Study*. Oxford, U.K.: Pergamon Press, 1980.

Chandler, M. J. "Egocentrism and Antisocial Behavior: The Assessment and Training of Social Perspective-Taking Skills." *Developmental Psychology* 9(1973):1–6.

Cleary, T. A. "Test Bias: Prediction of Grades of Negro and White Students in Integrated Colleges." *Journal of Educational Measurement* 5(1968):115–124.

Cooper, C. R., and Odell, L., eds. *Evaluating Writing: Describing, Measuring, Judging*. Urbana, IL: National Council of Teachers of English, 1977.

Cronbach, L. J., and Gleser, C. C. *Psychological Tests and Personnel Decisions*. 2d ed. Urbana, IL: University of Illinois Press, 1965.

Cronbach, L. J. "Test Validation." In *Educational Measurement*. 2d ed., edited by R. L. Thorndike. Washington, DC: American Council on Education, 1971.

Crowhurst, M. "On the Misinterpretation of Syntactic Complexity Data." *English Education* 11(1979):91–97.

Crowhurst, M. "The Effect of Audience and Mode of Discourse on the Syntactic Complexity of the Writing of Sixth and Tenth Graders." Unpublished dissertation, University of Minnesota, 1977.

Davitz, J. R., and Mattis, S. "The Communication of Emotional Meaning of Metaphor." In J. R. Davitz (Ed.), *The Communication of Emotional Meaning*. New York: McGraw Hill, 1964.

De La Zerda, N., and Hopper R. "Employment Interviewers' Reactions to Mexican American Speech." *Communication Monographs* 46(1979):126–134.

Dickson, W. P., and Patterson, J. H. "Evaluating Referential Communication Games for Teaching Speaking and Listening Skills." *Communication Education* 30(1979):11–21.

Diederich, P. B. *Measuring Growth in English*. Urbana, IL: National Council of Teachers of English, 1974.

Educational Testing Service. *Say and Tell Circus A and B User's Guide*. Menlo Park, CA: Addison-Wesley, 1979.

Fagan, W.; Cooper, C.; and Jensen, J. *Measures for Research and Evaluation in the English Language Arts*. Urbana, IL: National Council of Teachers of English, 1975.

Follard, D., and Robertson, D. "Towards Objectivity in Group Oral Testing." *English Language Teaching Journal* 30(1976):156–167.

Gary Community Schools Corporation. *Oral Proficiency Program: Secondary Schools 1977–1978*. Gary, IN: author, 1977–1978.

Gay, G., and Abrahams, R. D. "Does the Pot Melt, Boil, or Brew? Black Children and White Assessment Procedures." *Journal of School Psychology* 11(1973):330–340.

Grommon, A. H., ed. *Reviews of Selected Published Tests in English*. Urbana, IL: National Council of Teachers of English, 1976.
Gruner, C. R., Logue, C. M., Freshley, D. V., and Huseman, R. C. *Speech Communication in Society*. Boston: Allyn and Bacon, 1977.
Hitchman, P. J. "CSE Tests in Spoken English." *Educational Research* 20(1968):218–225.
Hitchman, P. J. *Examining Oral English in Schools*. London: Methuen, 1966.
Hudson, S., and Veal, L. R. *Assessing Writing Competence Through Writing Samples* (Studies in Language Education Report No. 35). Athens, GA: University of Georgia Department of Language Education, 1979.
Hurt, H. T., and Preiss, R. "Silence Isn't Necessarily Golden: Communication Apprehension, Desired Social Choice, and Academic Success Among Middle-school Students." *Human Communication Research* 4(1978):315–328.
Johnson, O. G. *Tests and Measurements in Child Development: Handbook II* (2 vols.). San Francisco: Jossey-Bass, 1976.
Kelly, C. M. "An Investigation of the Construct Validity of Two Commercially Published Listening Tests." *Speech Monographs* 32(1965):139–143.
Kendall, K. E. "Do Real People Ever Give Speeches?" *Central States Speech Journal* 25(1974):233–235.
Lange, D. L., and Clifford, R. T. *Testing in Foreign Languages, ESL, and Bilingual Education, 1966–1979*. (Language in Education: Theory and Practice, No. 24). Washington, DC: Center for Applied Linguistics/ERIC, 1980.
Larson, C.; Backlund, P.; Redmond, M.; and Barbour A. *Assessing Functional Communication*. Falls Church, VA: Speech Communication Association, 1978.
Larson, C. E. "Problems in Assessing Functional Communication." *Communication Education* 27(1978):304–309.
Loban, W. *Language Development: Kindergarten Through Grade 12*. (Research Report No. 18). Urbana: National Council of Teachers of English, 1976.
Lundsteen, S. W. *Listening: Its Impact on Reading and the Language Arts*. Urbana: ERIC and National Council of Teachers of English, 1979.
McCaleb, J. L. "Measuring Oral Communication." *English Education* 68(1979):41–47.
McCaleb, J., and Korman, D. L. "A Measure of Communication Development." *English Journal* 67(1978):41–45.
McCroskey, J. "Oral Communication Apprehension: A Summary of Recent Theory and Research." *Human Communication Research* 4(1977):78–96.
McCroskey, J. C., and Daly, J. A. "Teachers' Expectations of the Communication Apprehensive Child in the Elementary School." *Human Communication Research* 3(1976):67–72.
McGlone, E. L. "Educational Measurement and Speech Communication Instruction." *Central States Speech Journal* 24(1973):241–246.
Maslow, A. H. *Motivation and Personality*. New York: Harper & Row, 1954.
Massachusetts Listening and Speaking Assessment. *State Listening Test (Secondary)*. Boston: author, 1980.
Massachusetts Listening and Speaking Assessment. *State Speaking Test (Secondary)*. Boston: author, n.d.
Mead, G. H. *Mind, Self, and Society*. Chicago: University of Chicago Press, 1934.
Mead, N. A. "The Development of an Instrument for Assessing Functional Communication Competence of Seventeen-Year-Olds." Unpublished dissertation, University of Denver, 1977.
Mecham, M. J., and Jones, J. D. *Utah Test of Language Development*. rev. ed. Salt Lake City, Utah: Communication Research Associates, 1978.
Mullen, K. A. "Direct Evaluation of Rater and Scale in Oral Interviews." *Language Learning* 28(1978):301–308.
Philips, S. U. "Acquisition of Rules for Appropriate Speech Usage." *Monograph Series on Language and Linguistics* (No. 23). Washington, DC: Georgetown University Press, 1970.
Piche, G. L.; Rubin, D. L.; and Turner, L. J. "Training for Referential Communication Accuracy in Writing." *Research in the Teaching of English* 14(1980):309–318.
Plattor, E.; Unruh, W. R.; Muir, L.; and Loose, K. D. *Test Development for Assessing Achievement in Listening and Speaking*. Edmonton, Alberta: Alberta Education, 1978.
Price, W. K. *The University of Wisconsin Speech Attainment Test*. Unpublished dissertation, University of Wisconsin, 1964.
Richard, D. D. "Communication Competency Assessment for Non-Native Speakers of English." Paper presented at the Annual Meeting of the Southern Speech Communication Association, Austin, Texas, April, 1981.

Richmond, V. P. "Communication Apprehension and Success in the Job Applicant Screening Process." Paper presented at the Annual Meeting of the International Communication Association, Berlin, 1977.

Ritter, E. M. " 'Back to Basics' and Accountability Issues in Secondary School Speech Education." *Communication Education* 27(1978):119–126.

Rosen, H. *An Investigation of the Effects of Differentiated Writing Assignments on the Performance of a Selected Group of 15/16-year-old Pupils*. Unpublished dissertation, University of London, 1969.

Rubin, D. L. "Psychometric and Pedagogical Validity in Large Scale Assessment of Oral Communication Skills." Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980a.

Rubin, D. L. "Status of Speech Communication Instruction in the Georgia's Secondary Schools." *English Counsellor*, Spring, 1980b; n.p.

Rubin, D. L. "Using Performance Rating Scales in Large Scale Assessments of Speaking Proficiency." In *Perspective on Oral Communication Assessment in the 80's*, edited by R. Stiggins. Portland, OR: Northwest Regional Educational Laboratory, 1981.

Scott, M. D.; Yates, M.; and Wheeless, L. R. "An Exploratory Investigation of the Effects of Communication Apprehension on Alternative Systems of Instruction." Paper presented at the Annual Meeting of the International Communication Association, Chicago, 1975.

Shure, M. B., and Spivack, G. *Problem-Solving Techniques in Childrearing*. San Francisco: Jossey-Bass, 1978.

Stiggins, R. J. "Potential Sources of Bias in Speaking and Listening Assessment." In *Perspectives on the Assessment of Speaking and Listening Skills for the 1980's*, edited by R. J. Stiggins. Portland, OR: Northwest Regional Educational Laboratory, 1981

Silvaroli, N. J.; Skinner, J. T.; and Maynes, J. O. *Oral Language Evaluation Teacher's Manual*. St. Paul, MN: EMC Corporation, 1977.

Smith-Elliot Listening Test. Needham, MA: Learning Dynamics, Inc., n.d.

Speech Communication Association. "Criteria for Evaluating Instruments and Procedures for Assessing Speaking and Listening." *SPECTRA* 15(1979):5.

Speech Communication Association. "Standards for Effective Oral Communication Programs." *SPECTRA* 15(1979):7.

Vermont Department of Education. *Basic Compentencies: A Manual of Information and Guidelines for Teachers and Administrators, Vermont Basic Competency Program*. Montpelier: author, 1977.

Wiemann, J. M. "Explication and Test of a Model of Communicative Competence." *Human Communication Research* 4(1977):195–213.

Wiemann, J. M., and Backlund, P. "Current Theory and Research in Communicative Competence." *Review of Educational Research* 50(1980):185–199.

Wilkinson, A. "The Testing of Oracy." In *Language Testing Symposium: A Psycholinguistic Approach*, edited by A. Davies. London: Oxford University Press, 1968.

Williams, F. *Explorations in the Linguistic Attitudes of Teachers*. Rowley, MA: Newbury House, 1976.